New Pathways towards Quantum-Encoded Data Analysis in Neutrino Physics

Jeff Lazar

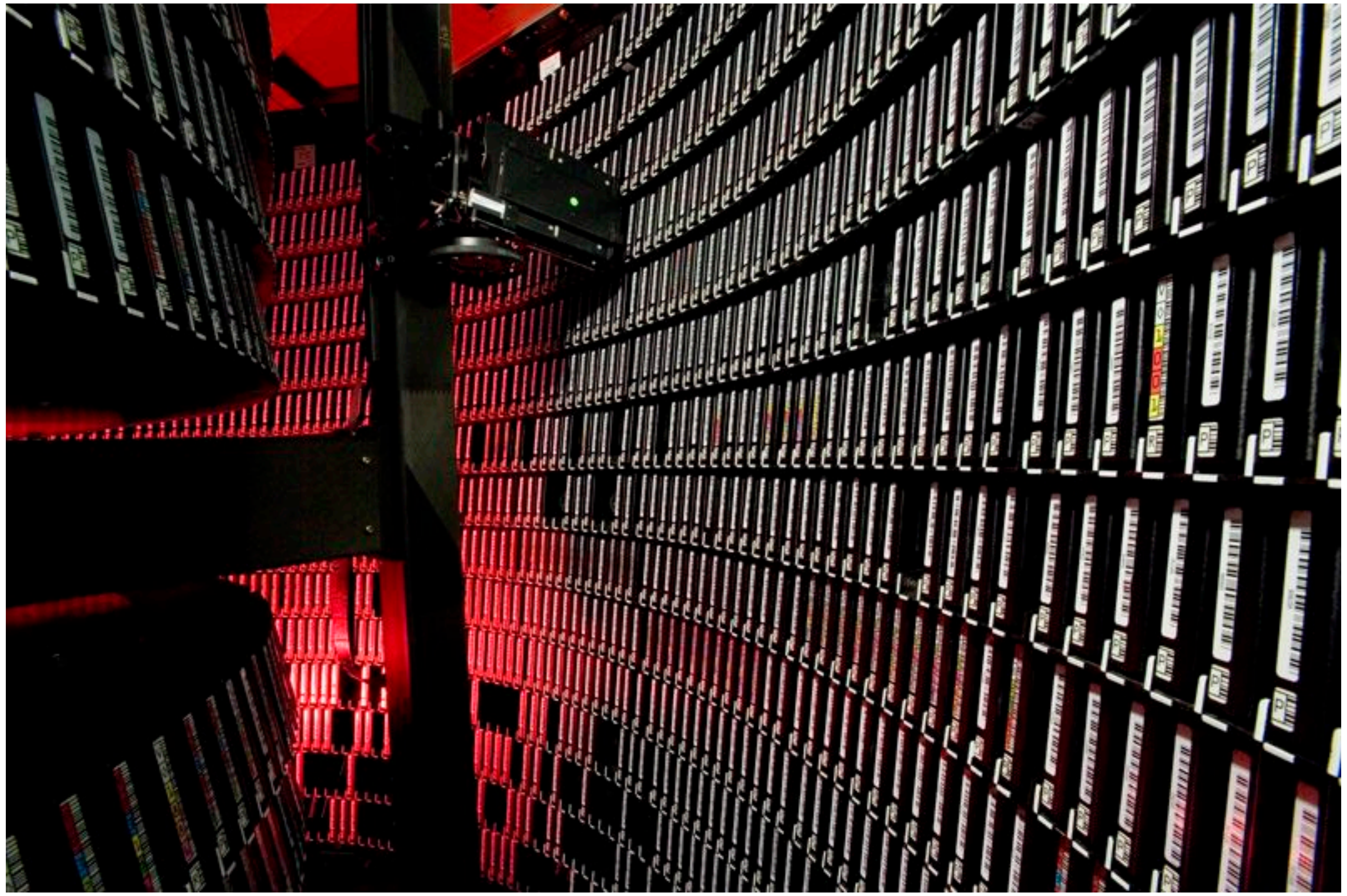
CERN QC Seminar

03 Oct., 2024

Geneva, Switzerland

Terabytes and Trouble

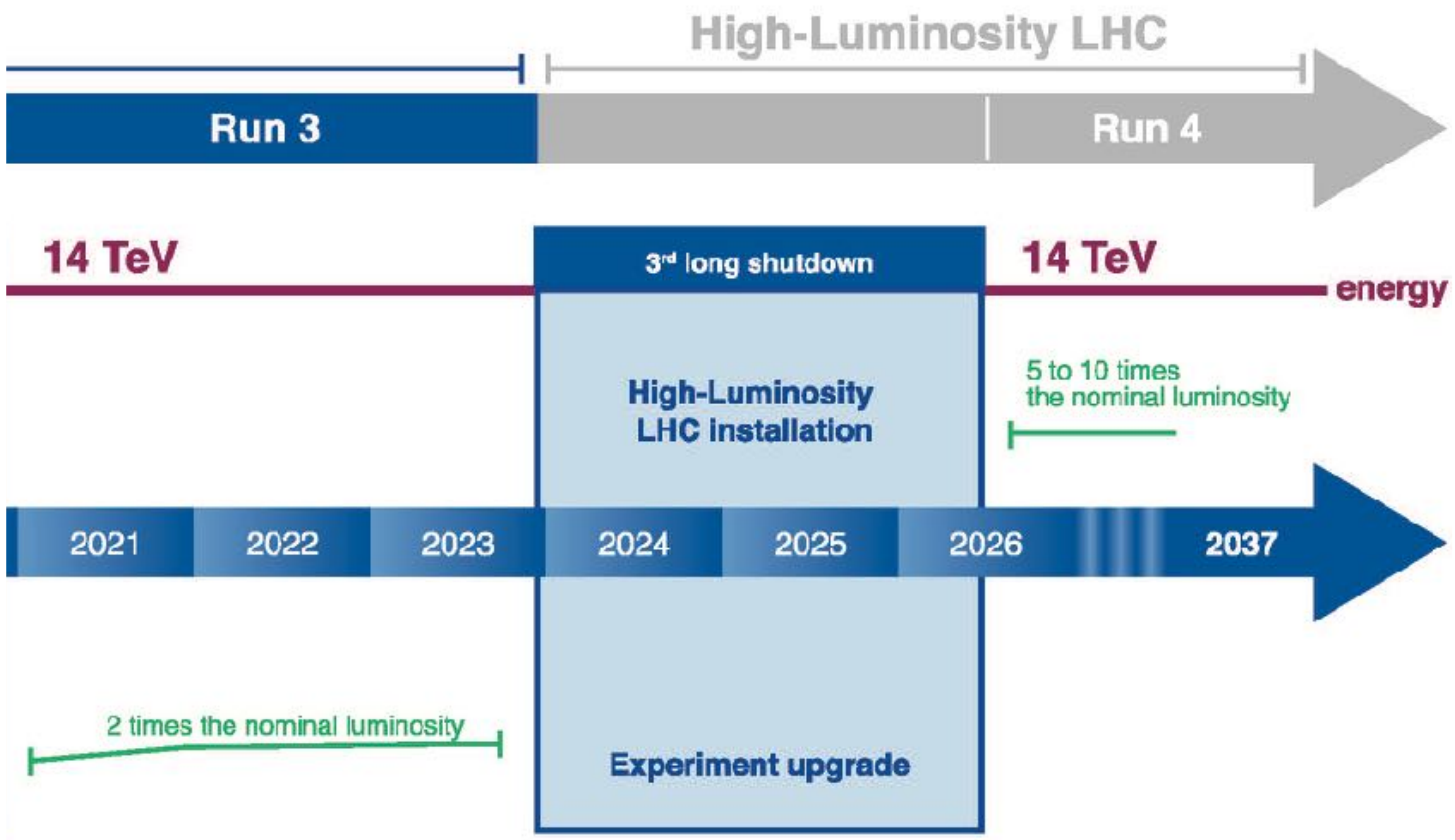
- Even after cuts, HEP experiments produce huge amounts of data !
- CERN produces > 300 TB of data per day
  - ~250 TB from LHC
  - ~70 TB from other experiments
- IceCube produces ~1 TB per day
- Sometimes multiple copies of this data must be stored

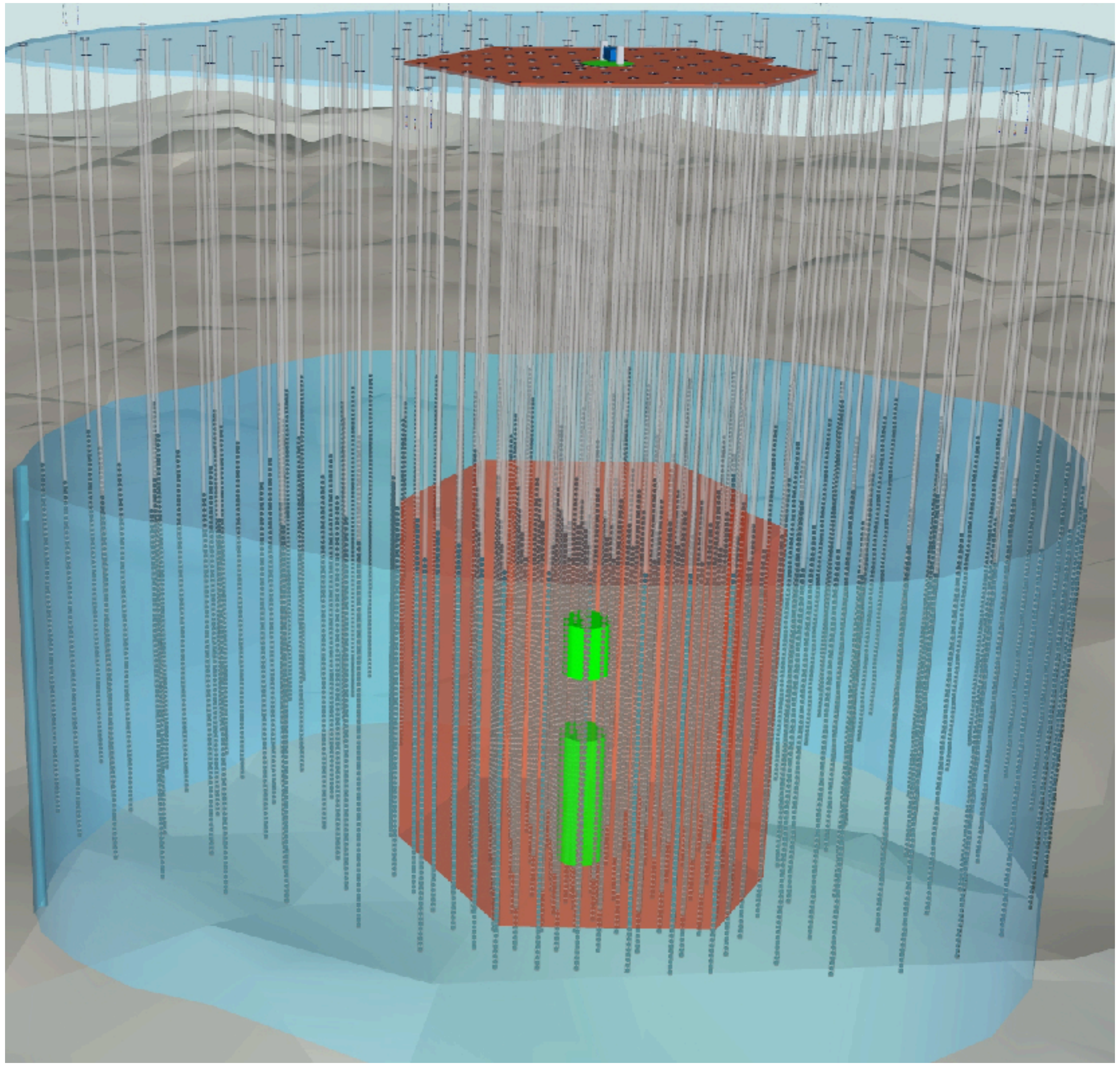


Data retrieval at Fermilab's Feynman Computing Center. A robotic arm retrieves and reads CMS data stored on hard drives at Fermilab
Photo: Reidar Hahn

Terabytes and Trouble

- Larger and more luminous experiments are on the horizon
- A growing problem





Outline

- **Encoding Information in Quantum Random Access Codes**
- Example application to neutrino telescope data
- Concluding remarks

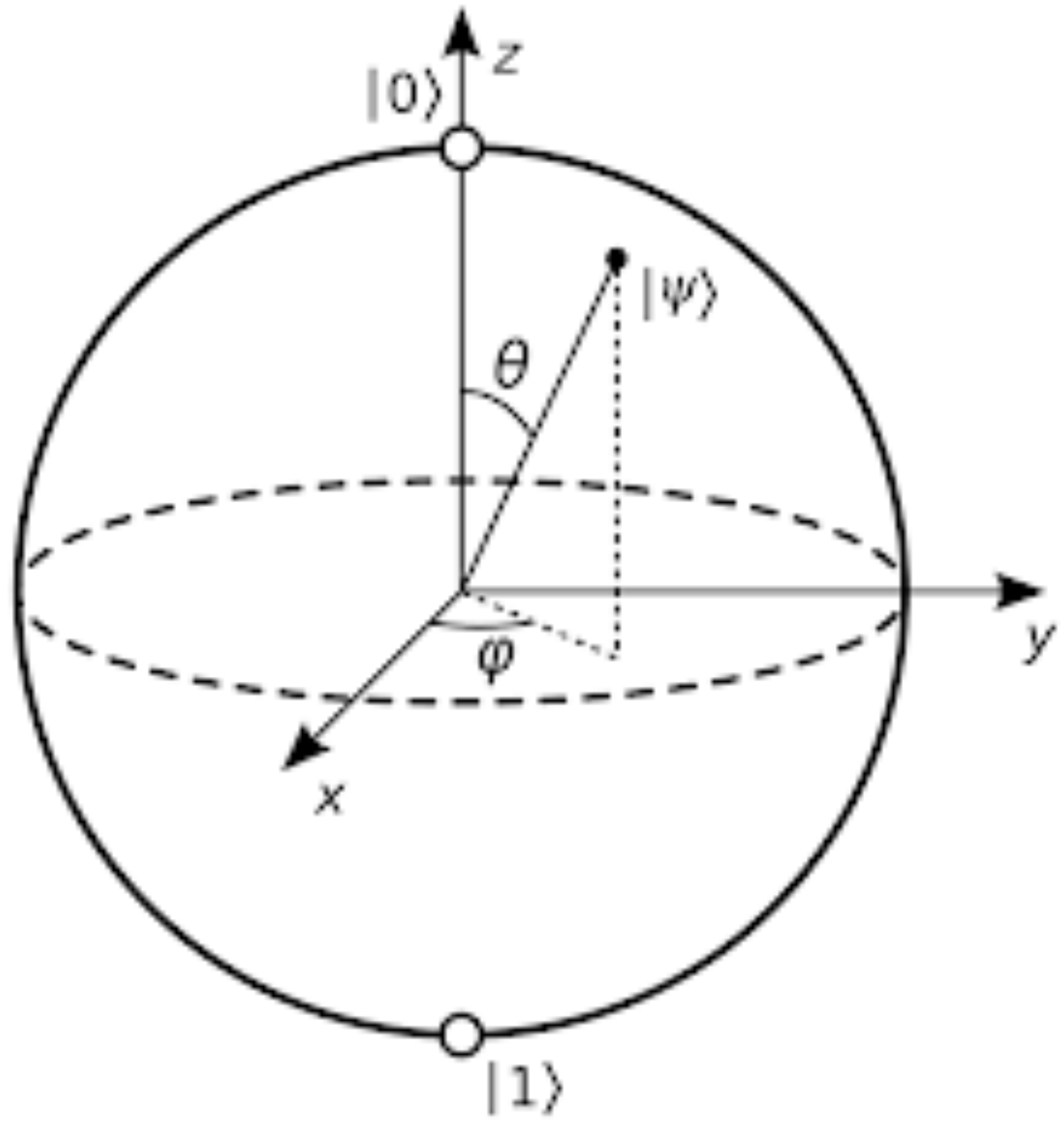
New Pathways in Neutrino Physics via Quantum-Encoded Data Analysis

Jeffrey Lazar,[1,2,*] Santiago Giner Olavarrieta,[2,†] Giancarlo Gatti,[3,4,5,‡] Carlos A. Argüelles,[2,§] and Mikel Sanz[3,4,6,7,¶]




Representing Data in Qubits

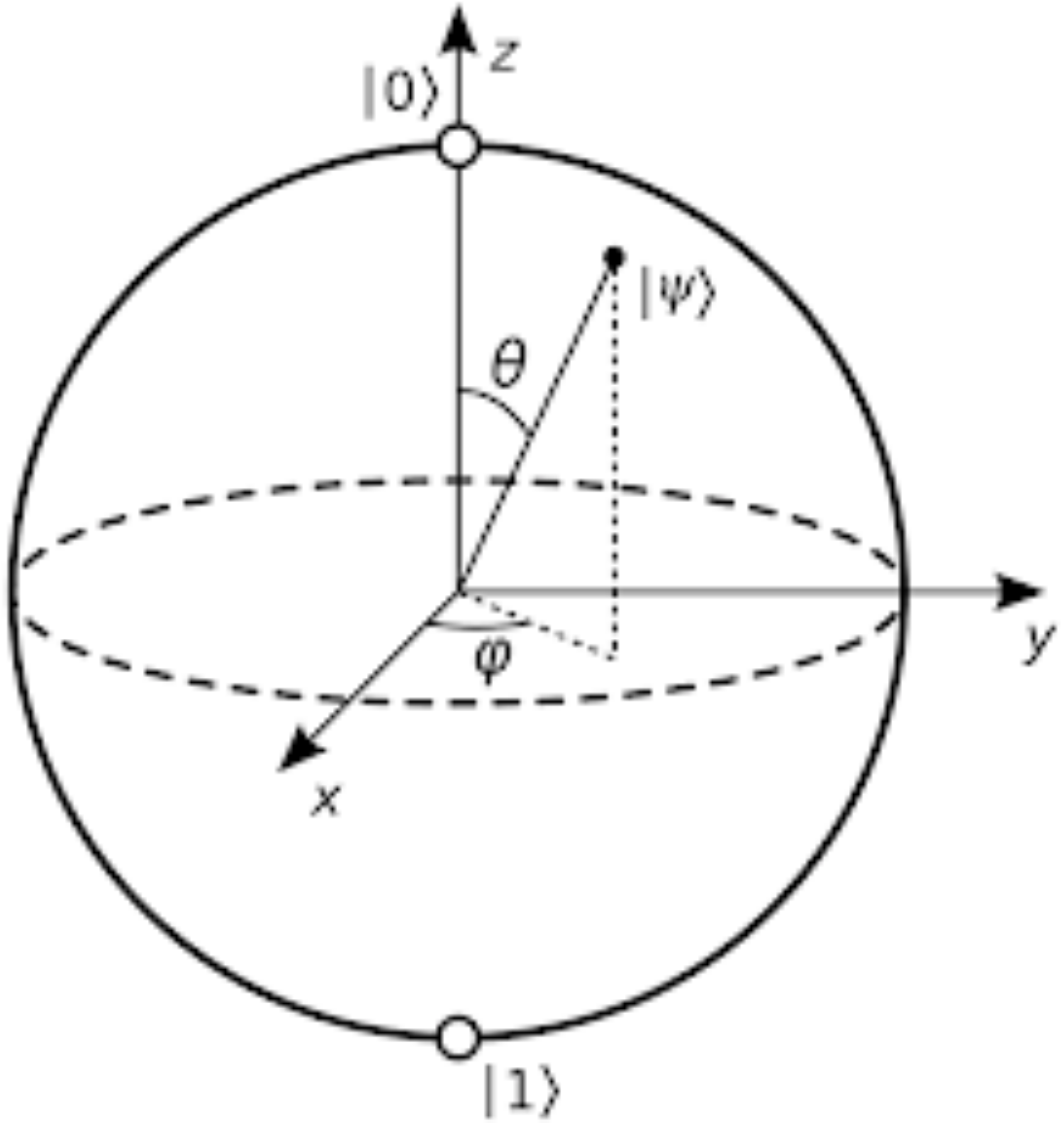
- Representing numerical data in qubits is non-trivial
- Angle encoding is used in much of the literature



Angle Encoding: An Analog Encoding

$$\phi, \theta = \arctan(d_1),\ \arctan(d_2)$$

- Embed data into angles by taking arctangent
- Only polar angle impacts expectation value
- Errors can dramatically affect encoded values
- Amount of data linear with number of qubits



Towards a Digital Encoding

- Qubits are two-level systems and so they are naturally suited to binary representations
- Naively idea would be to encode binary numbers
- Introduce binary operator $\hat{b}_z$ with eigenvalues 0 and 1

$$\hat{b}_z = \frac{1}{2}\left[\frac{2}{\hbar}\hat{z} + 1\right]$$

$$\hat{b}_z \mid \pm \rangle = \frac{1}{2}\left[\pm 1 + 1\right] \mid \pm \rangle$$

Towards a Digital Encoding

- In an n-qubit system, you could encode n bits of data
- Not great, but maybe there's something here…

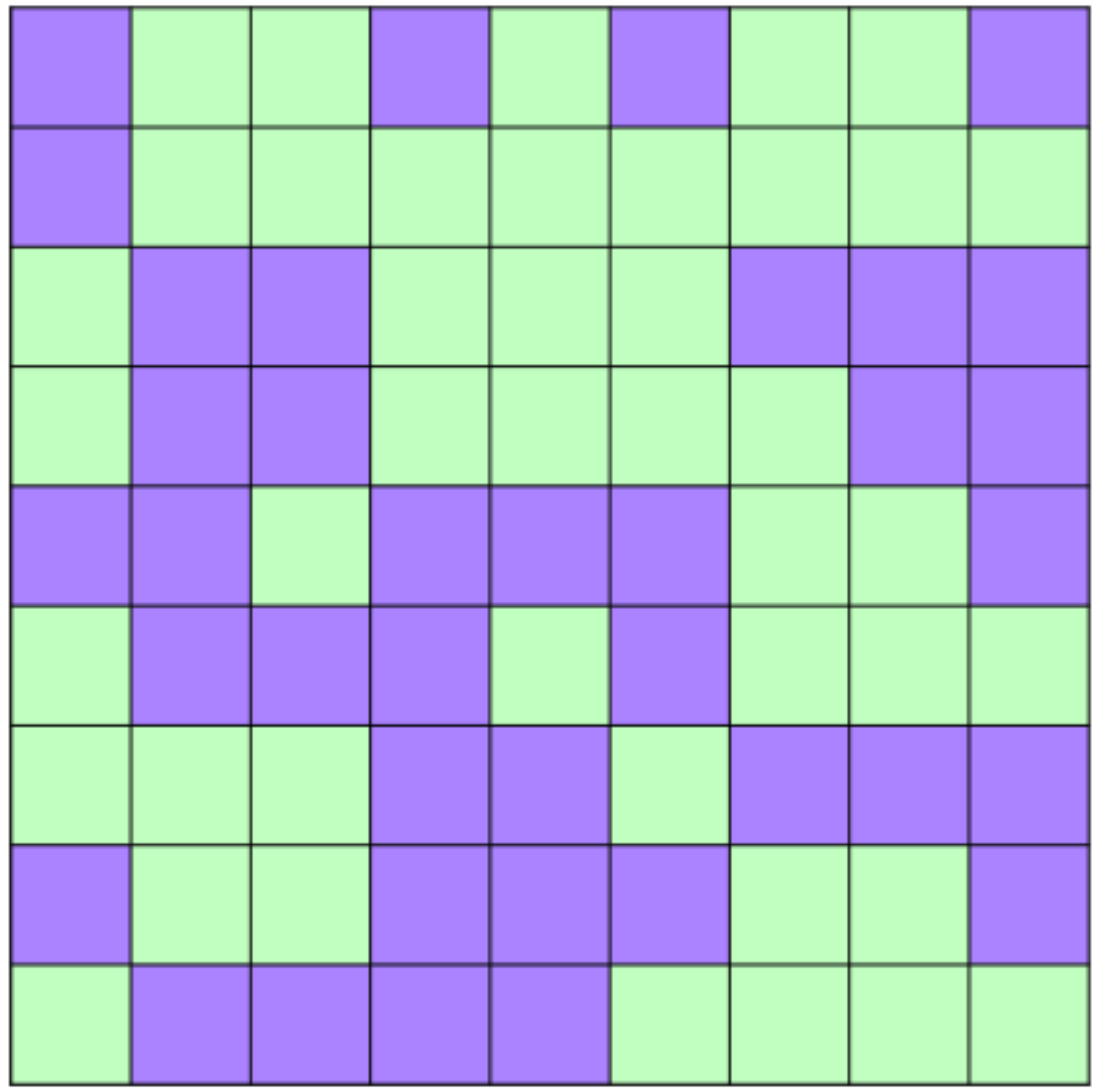
$$\hat{b}_z \left| q_0 \right\rangle = 0 \left| q_0 \right\rangle \qquad \hat{b}_z \left| q_1 \right\rangle = 1 \left| q_1 \right\rangle \qquad \hat{b}_z \left| q_2 \right\rangle = 1 \left| q_2 \right\rangle \qquad \hat{b}_z \left| q_3 \right\rangle = 0 \left| q_3 \right\rangle$$

0	1	1	0

Thinking about State Parity

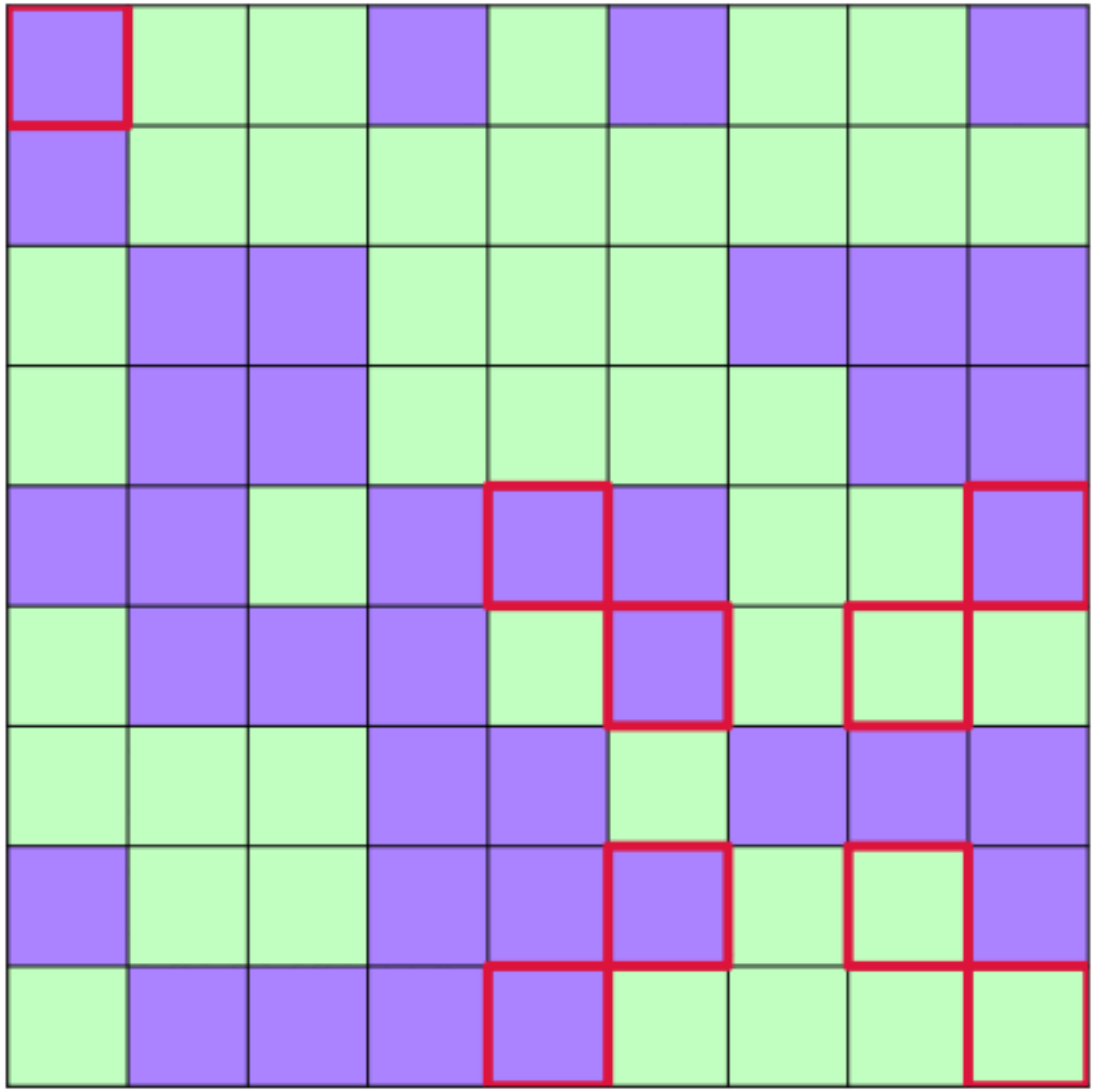
- Since the spin of any individual qubit is a binary outcome, the product over spins will also be a binary outcome
- We can now define a binary parity operator (PO), $\hat{b}^p_{\beta_0\beta_1\beta_2\beta_3}$, in a similar vein, with $\beta_i \in [z, x, y]$
- We map each classical bit to one of the 3^n POs, we have a lot of space

$$\hat{b}^p_{zzzz} = \frac{1}{2}\left[\left(\frac{2}{\hbar}\right)^4 \hat{z}_0\hat{z}_1\hat{z}_2\hat{z}_3 + 1\right]$$



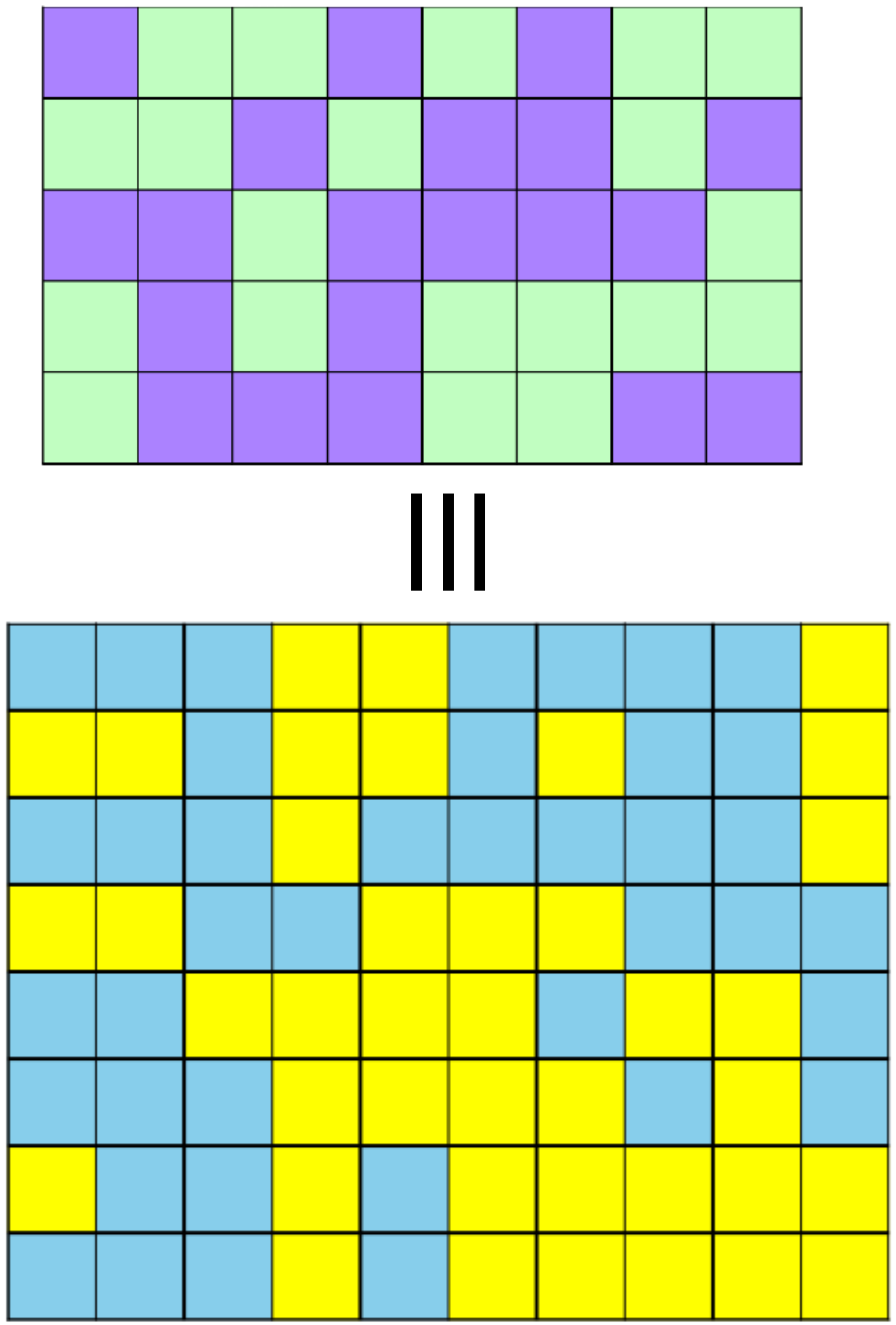
Parity and Quantum Random Access Codes

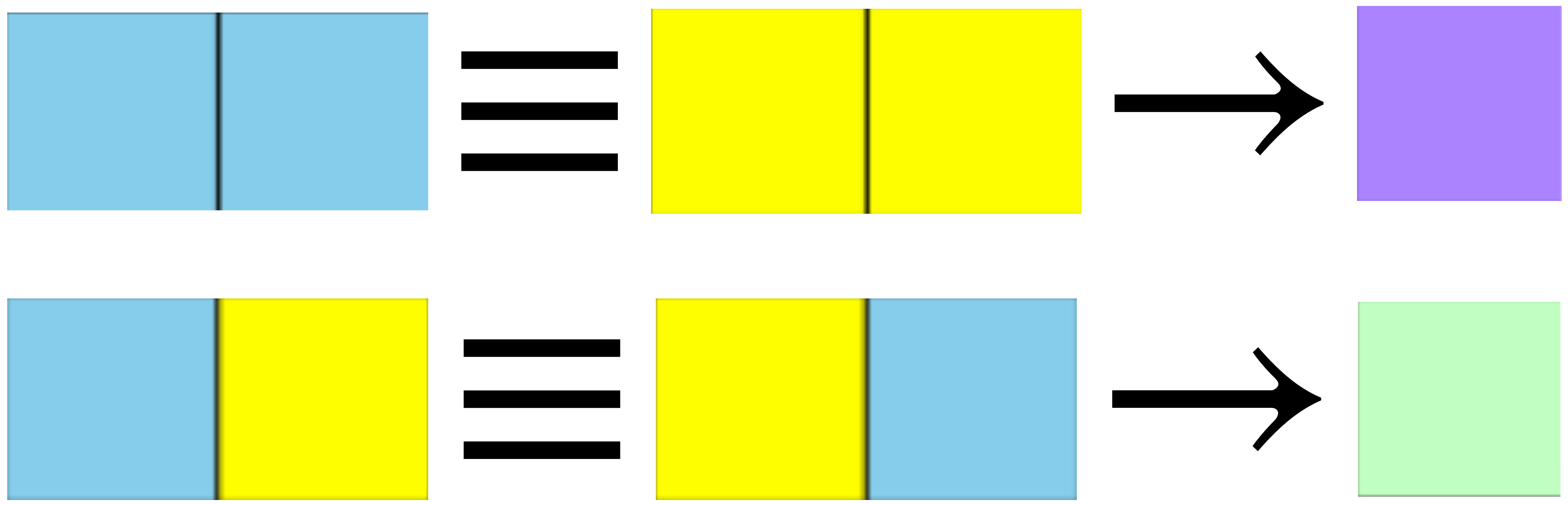
- Each n-qubit system will allow us to read a fraction of the total information
- Which portion of the information is determined by commutation relations
- But wait…
- There is no guarantee that a particular bit string will not conflict with allowed states



Going Back to Go Forward

- Instead, let's assign each classical bit to a pair of POs via XOR
- Encoding space halved, but gained exponential—$2^{\left(3^{n_{\text{qubit}}}-1\right)/2}$—freedom in ***representation***





Complete Sets of Commuting Observables

$|0\rangle$

$|0\rangle$

$|0\rangle$

$|0\rangle$

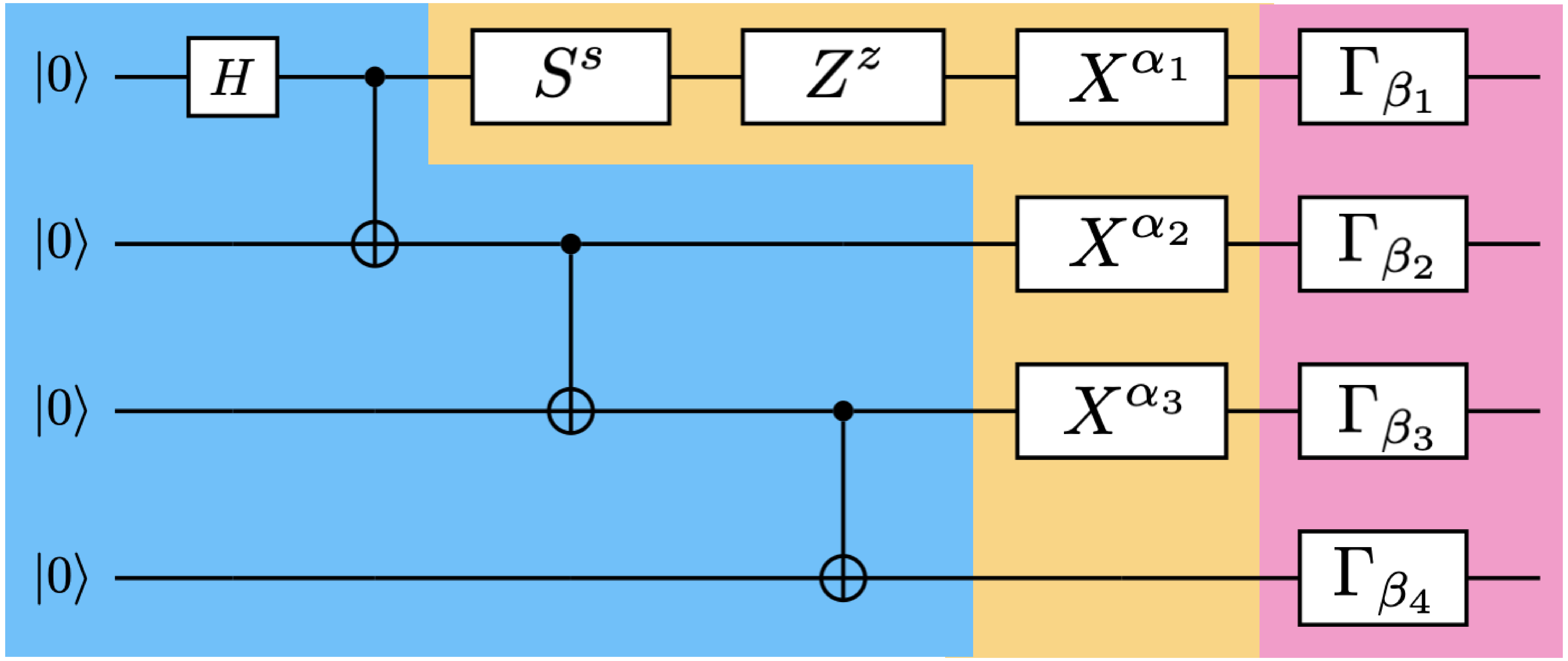
$$S^{\star}(\theta_s) = \begin{cases} I & \theta_s = 0 \\ S & \theta_s = 1 \end{cases} \qquad Z^{\star}(\theta_z) = \begin{cases} I & \theta_s = 0 \\ Z & \theta_s = 1 \end{cases}$$

$$X^{\star}(\alpha_x) = \begin{cases} I & \theta_s = 0 \\ X & \theta_s = 1 \end{cases} \qquad \Gamma(\beta) = \begin{cases} I & \beta = 0 \\ HZS & \beta = 1 \\ SH & \beta = 2 \end{cases}$$

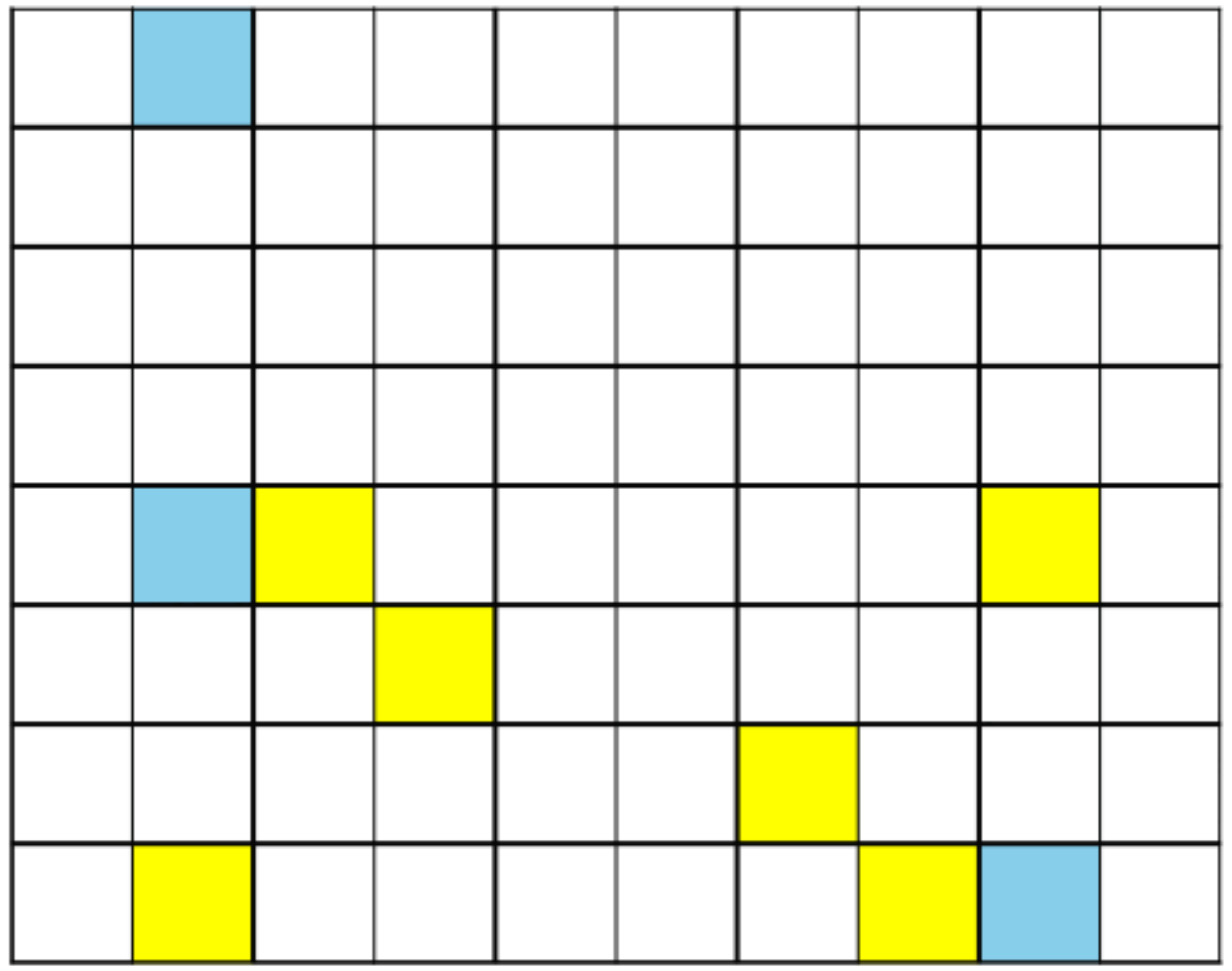
$$|0000\rangle \to \frac{1}{\sqrt{2}}\left[|0000\rangle + |1111\rangle\right] \to \frac{1}{\sqrt{2}}\left[|0000\rangle + (-1)^{\theta_z}|1111\rangle\right] \to \frac{1}{\sqrt{2}}\left[|0000\rangle + (i)^{\theta_s+2\theta_z}|1111\rangle\right]$$

$$\overset{\text{E.g.}}{\to} \frac{1}{\sqrt{2}}\left[|0110\rangle + (i)^{\theta_s+2\theta_z}|1001\rangle\right] \overset{\text{E.g.}}{\to} \frac{1}{\sqrt{2}}\left[|LR1+\rangle + (i)^{\theta_s+2\theta_z}|RL0-\rangle\right]$$

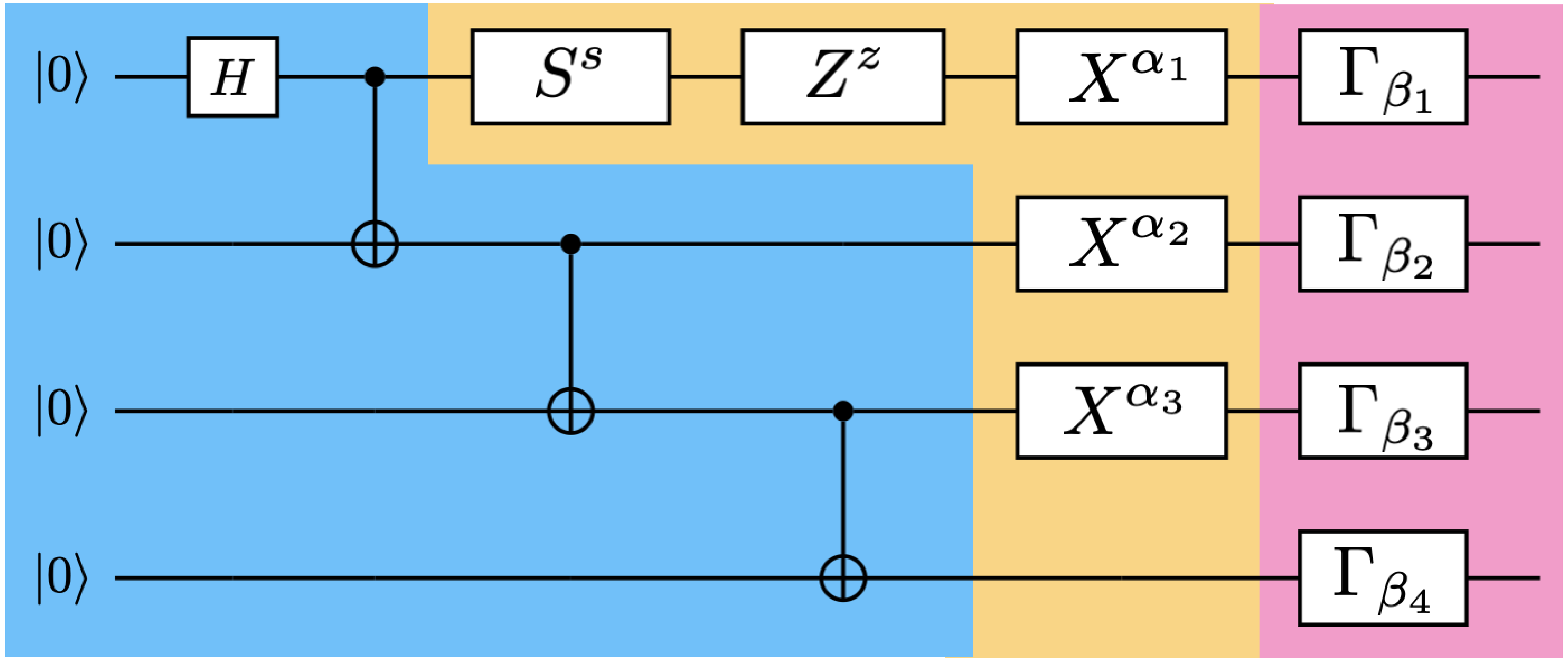
Turning the Knobs on CSCO Eigenstates



Changing α_i



CSCO Eigenstates by the Numbers

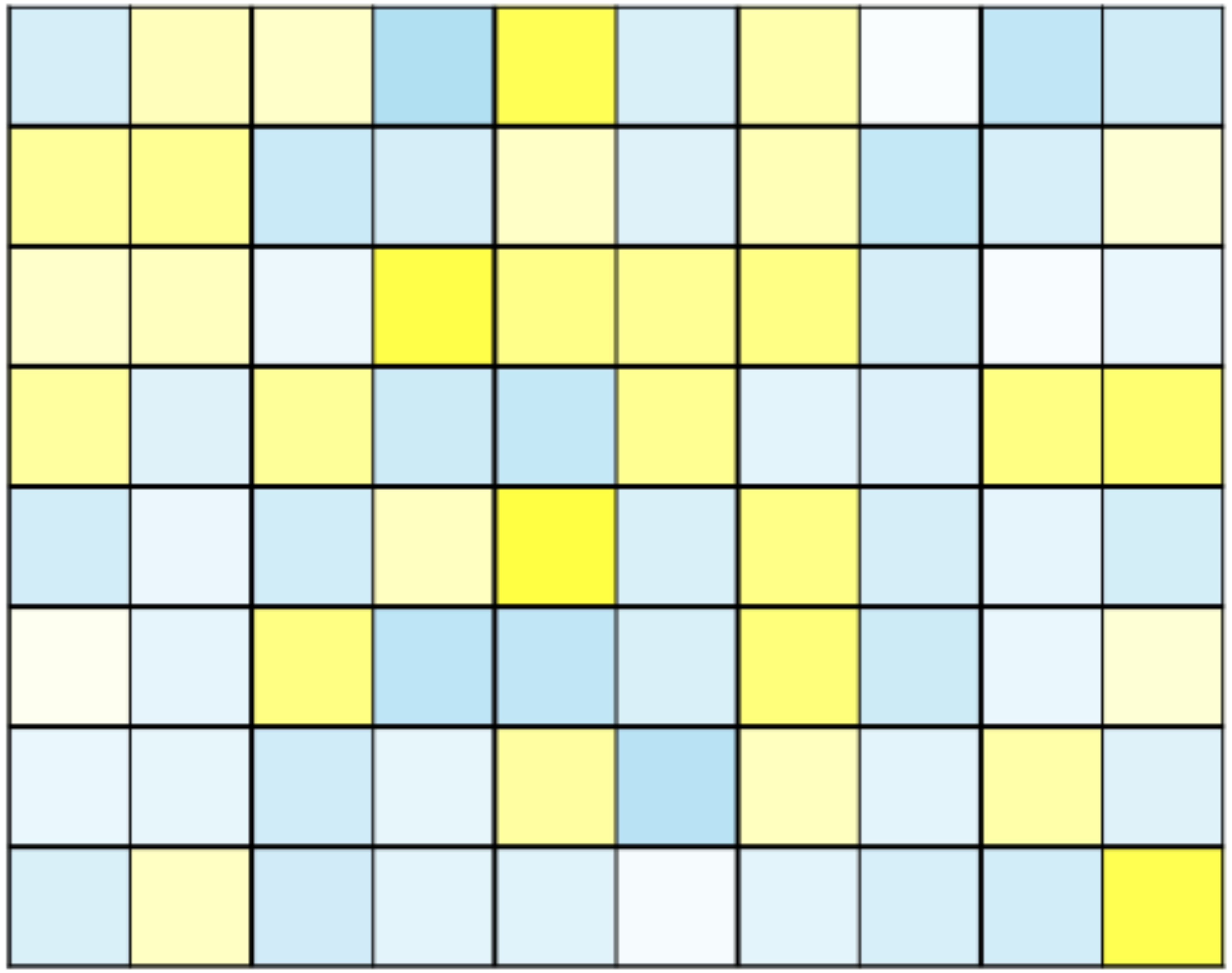


- There are 2×3^n CSCOs and each has 2^n allowed eigenstates
- Since each state has information on $2^{n-1} + 1$ observables, we have $\sim 12^n$ eigenvalues to sift through…
- Symmetries between different β values allow us to bring this to 4^n

Painting by Number

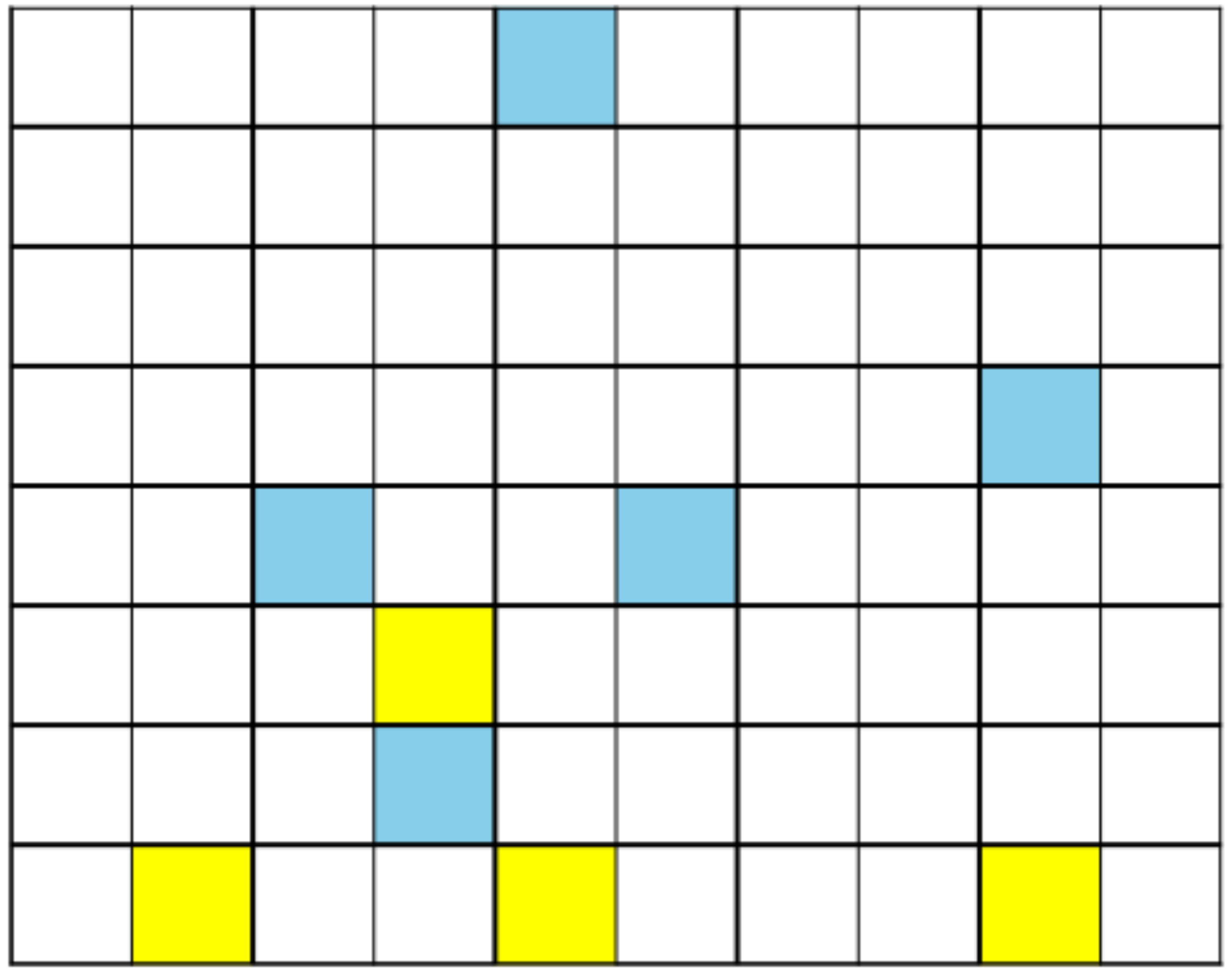
$$b_i = \text{round}\left[\text{bit}\left(\frac{1}{|\{|\phi\rangle\}|}\sum_j \langle\phi_j|\,\mathcal{O}(i)\,|\phi_j\rangle\right)\right]$$

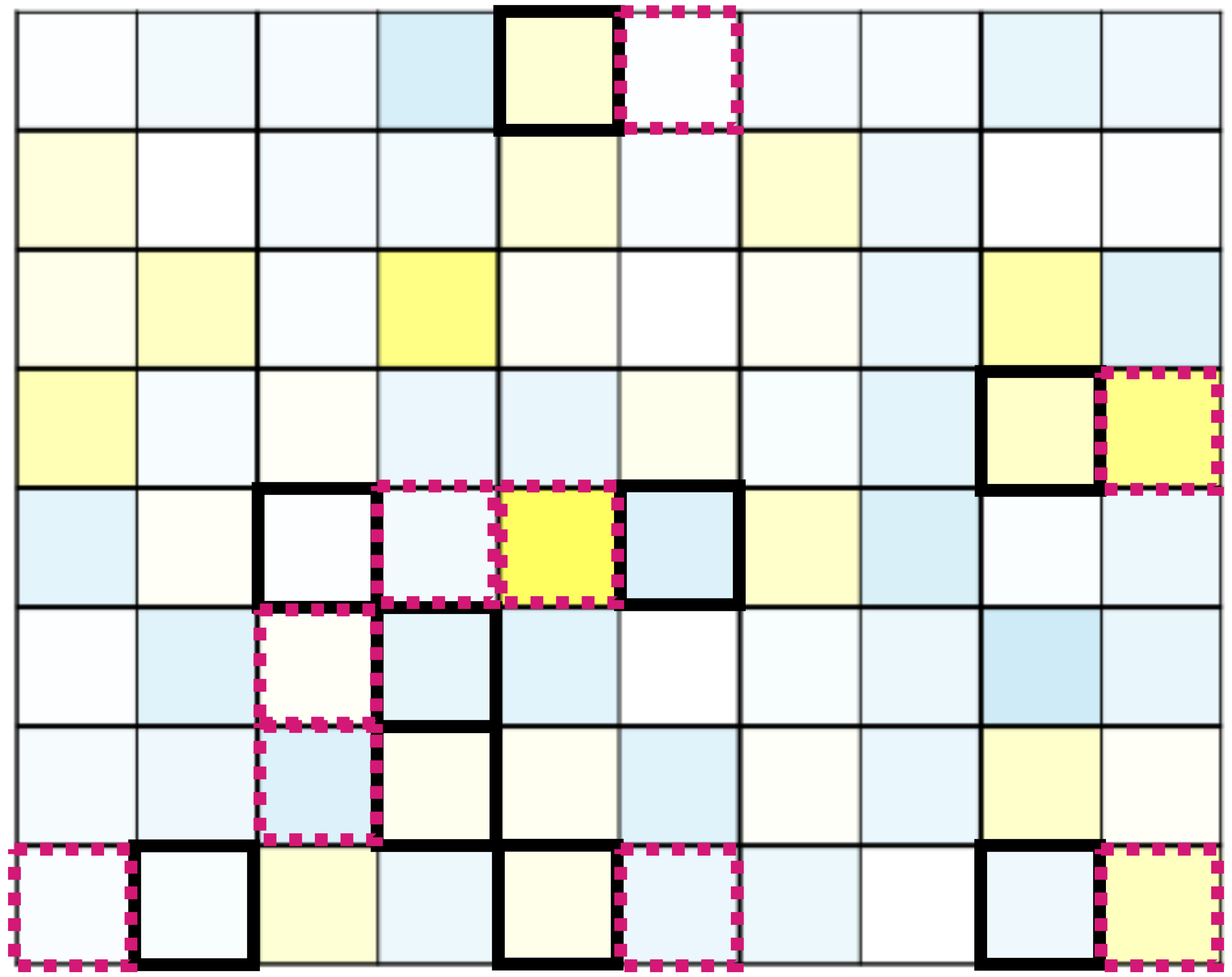
- Randomly select a number of states and take the average over all relevant CSCOs
- And then optimize those states
- Details of optimization are highly technical and somewhat varied



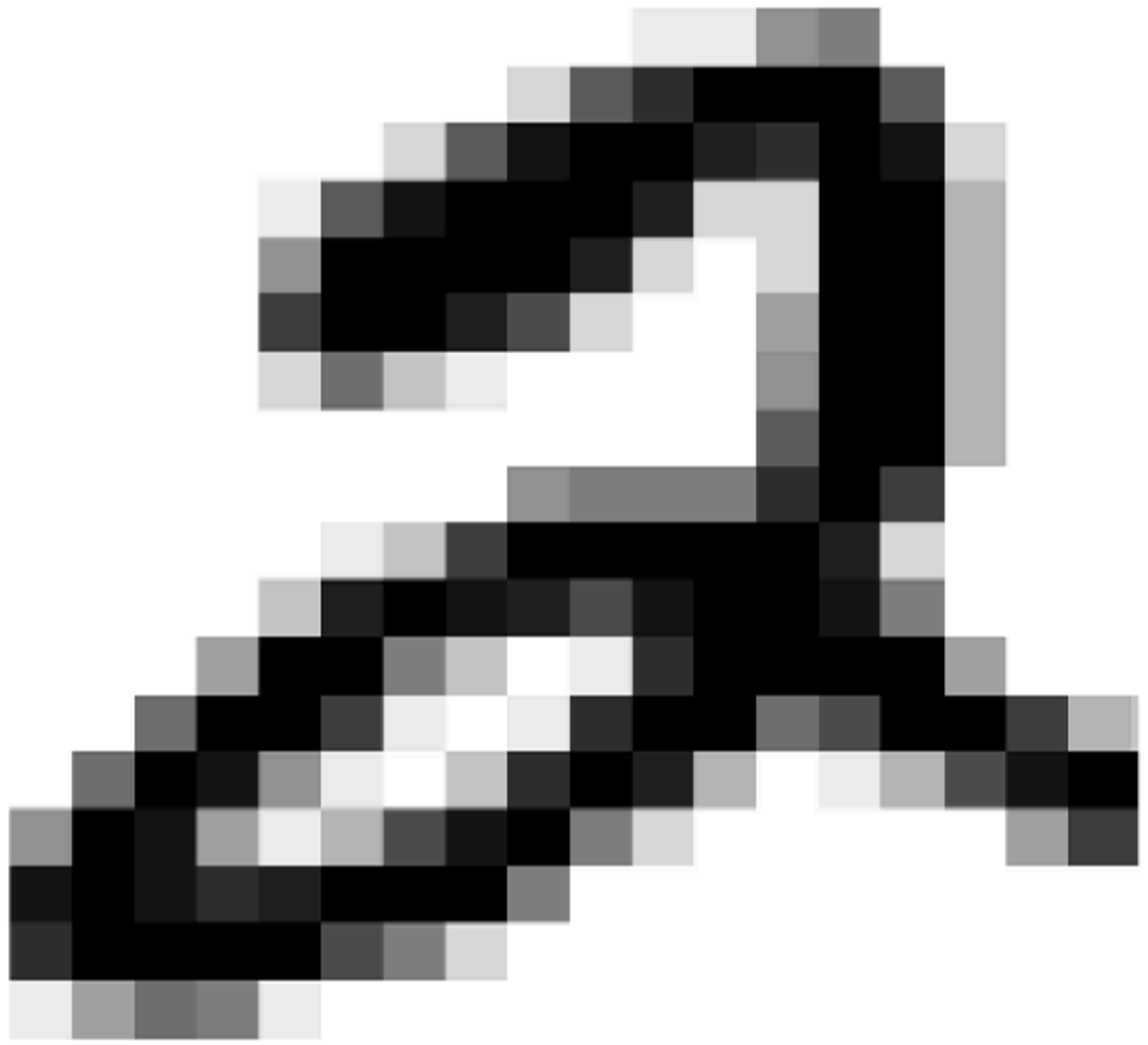
Optimization Scheme

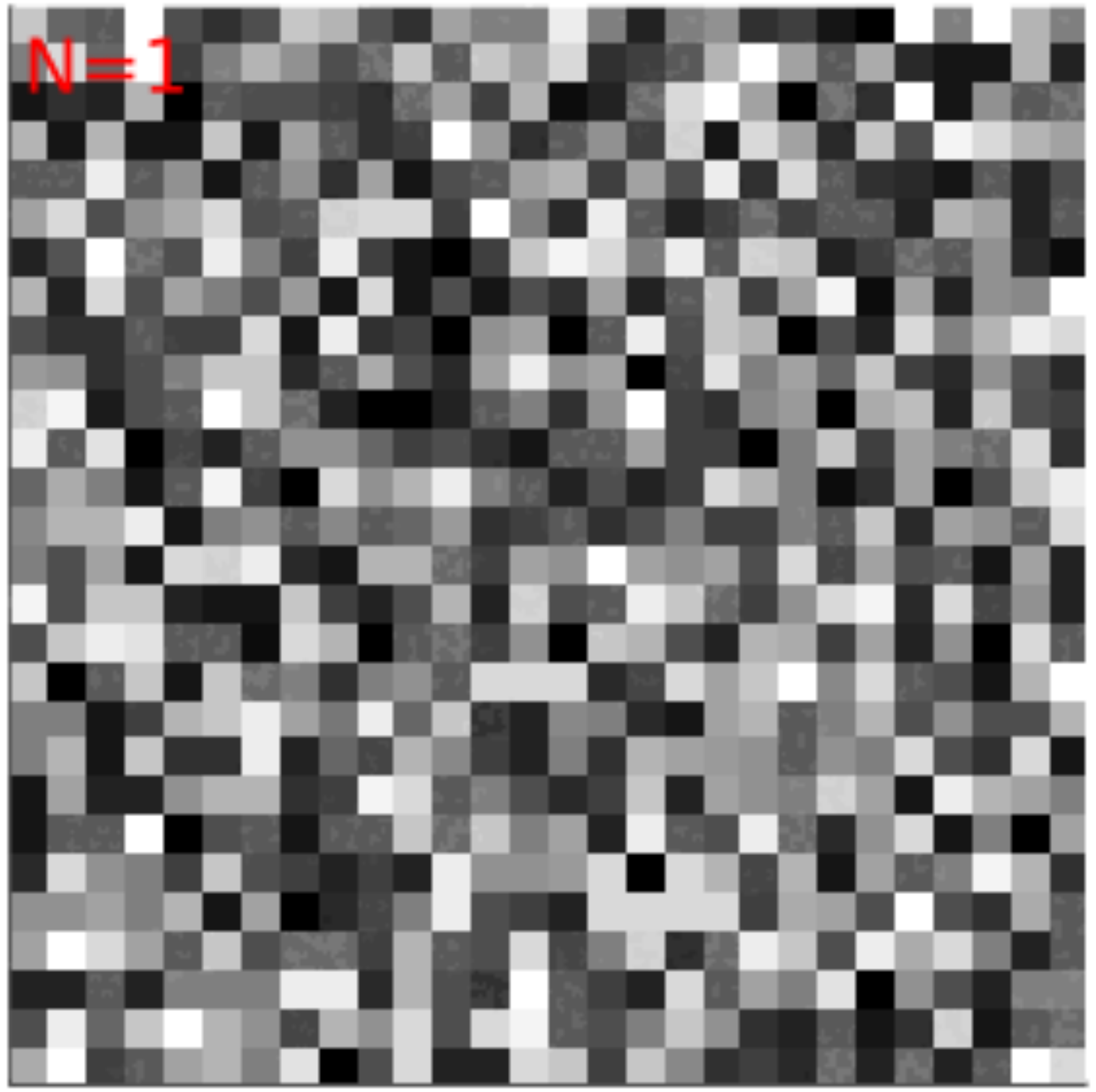
1. While convergence criteria not met
   - Score states based on whether they move the corresponding pair in the right direction
   - Preferentially select low-scoring state(s) to pick a new, better set of θ_z and α_i
2. Replace state(s) with new states that cover unbiased states, go to step 1





A Slightly Less Abstract Example





Can it compress ?

Maybe.

In this model, with redundancy r we need

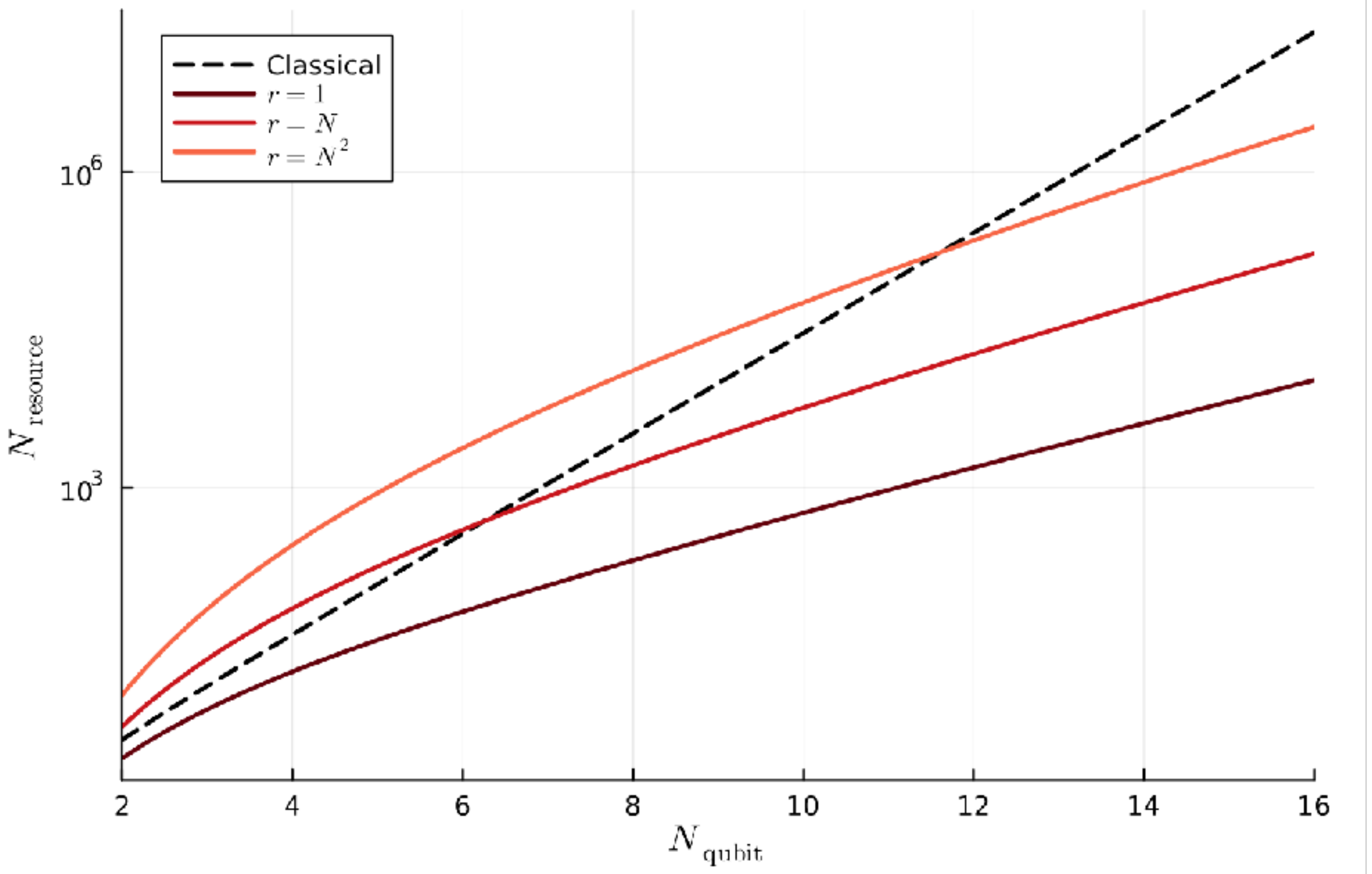
$$r \times n \times \frac{3^n - 1}{2 \times (2^{n-1} + 1)} \sim r \times n \left(\frac{3}{2}\right)^n$$

two-state systems to represent

$$\frac{3^{n-1} - 1}{2}$$

classical bits.

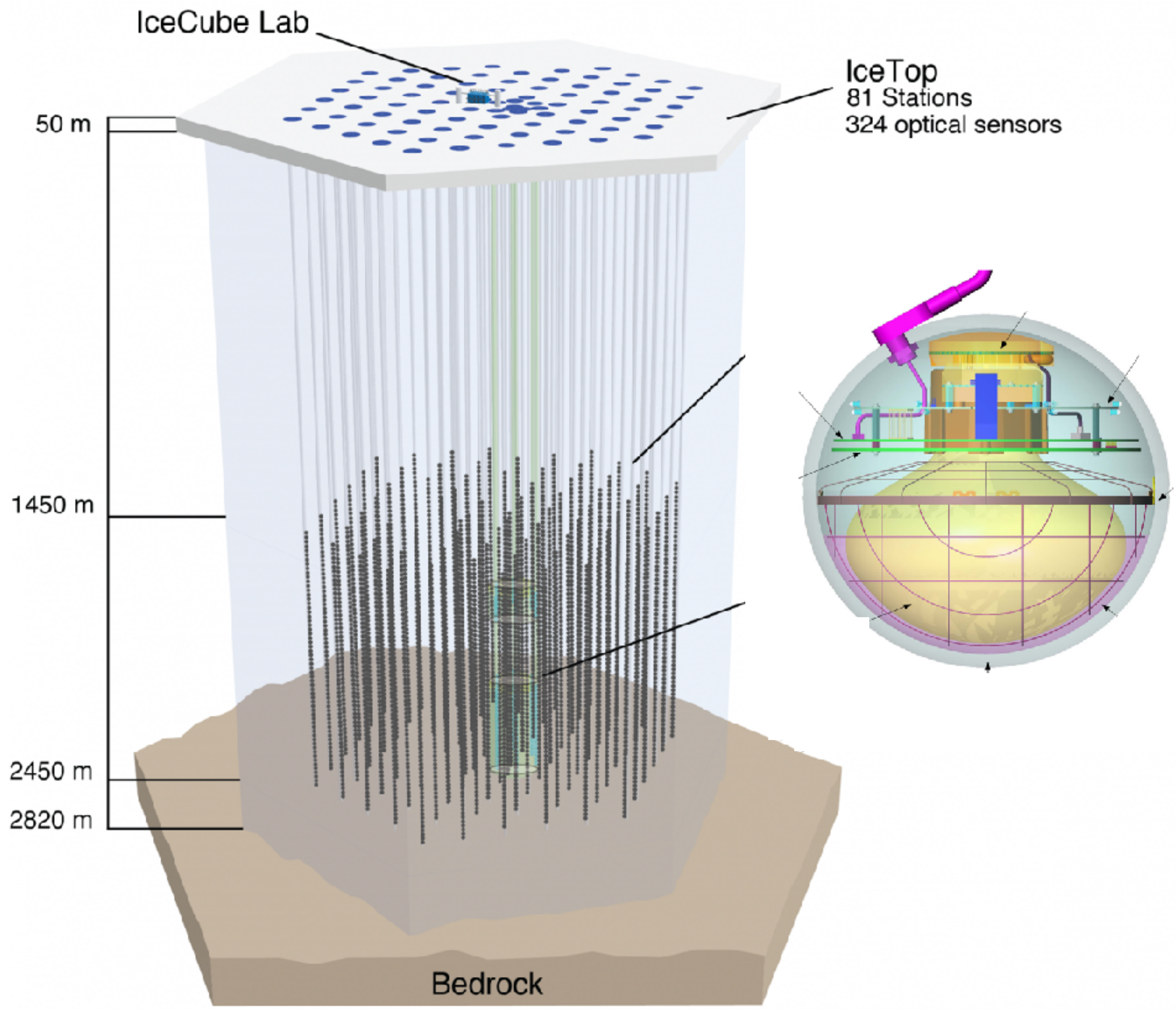
If r is polynomial, we will achieve compression



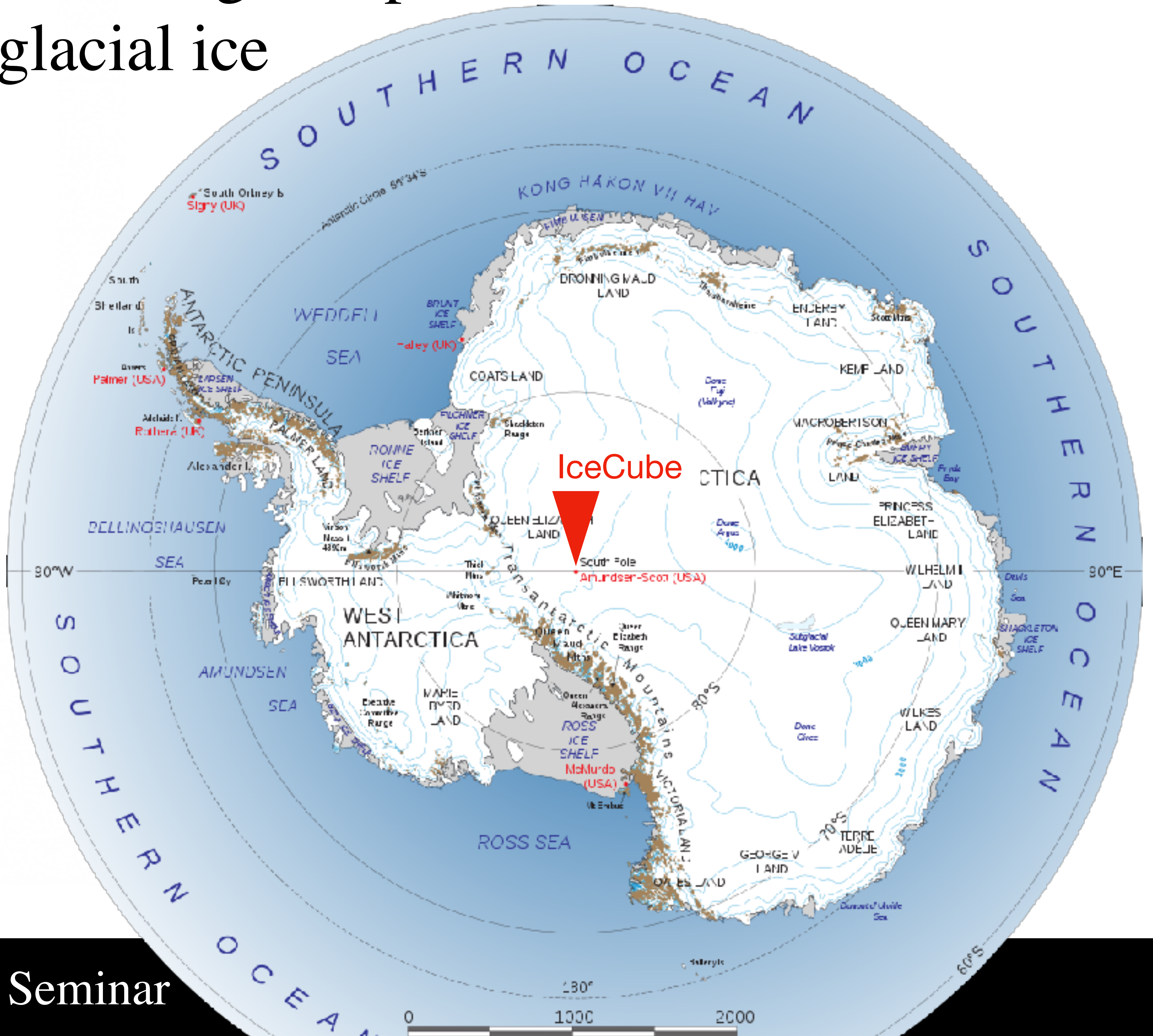
Outline

- Encoding Information in Quantum Random Access Codes
- **Example application to neutrino telescope data**
- Concluding remarks

IceCube Neutrino Observatory

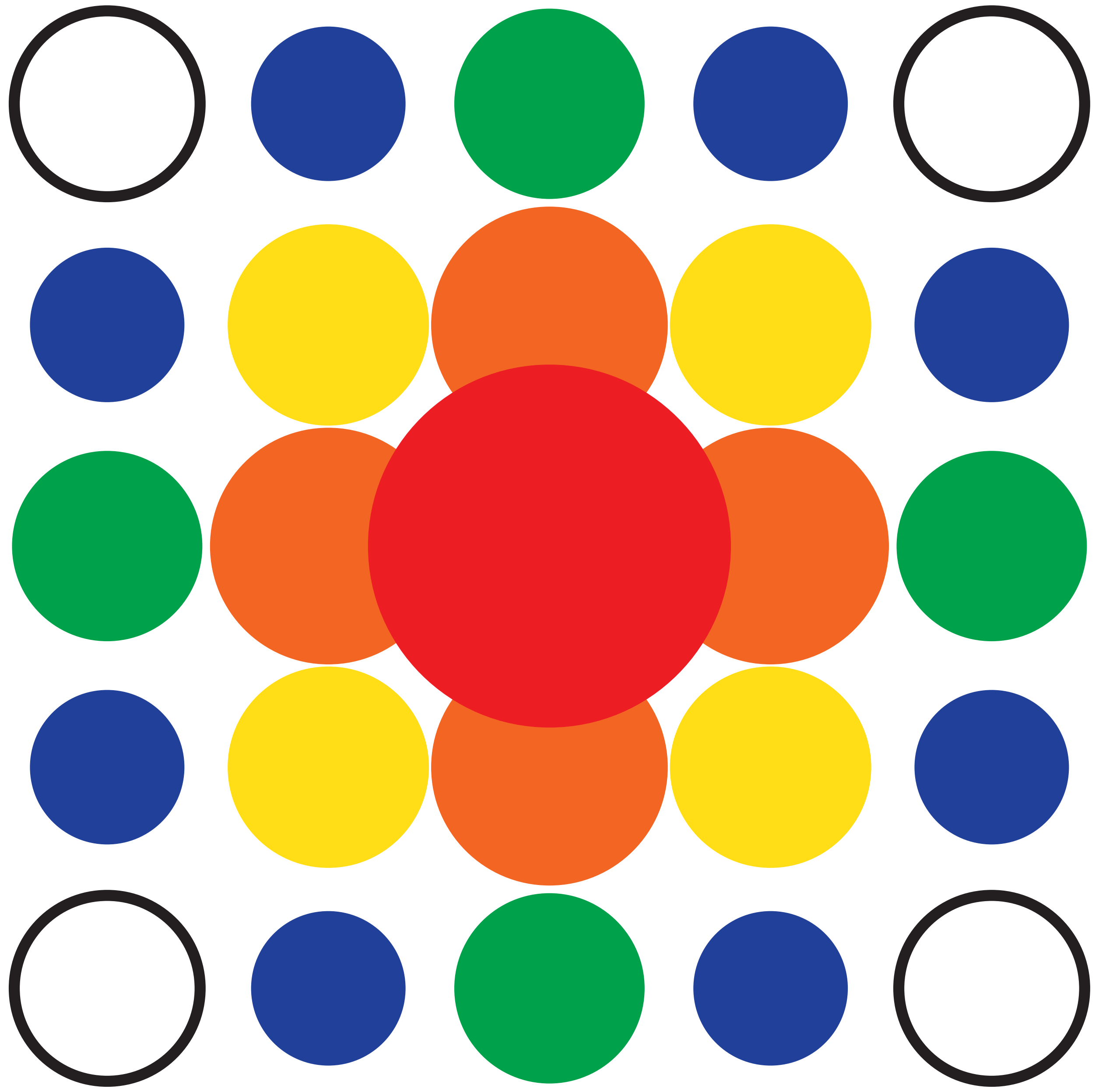


- IceCube Neutrino Observatory is largest operating neutrino telescope
- 5,160 digital optical modules in 1 km^3 of glacial ice



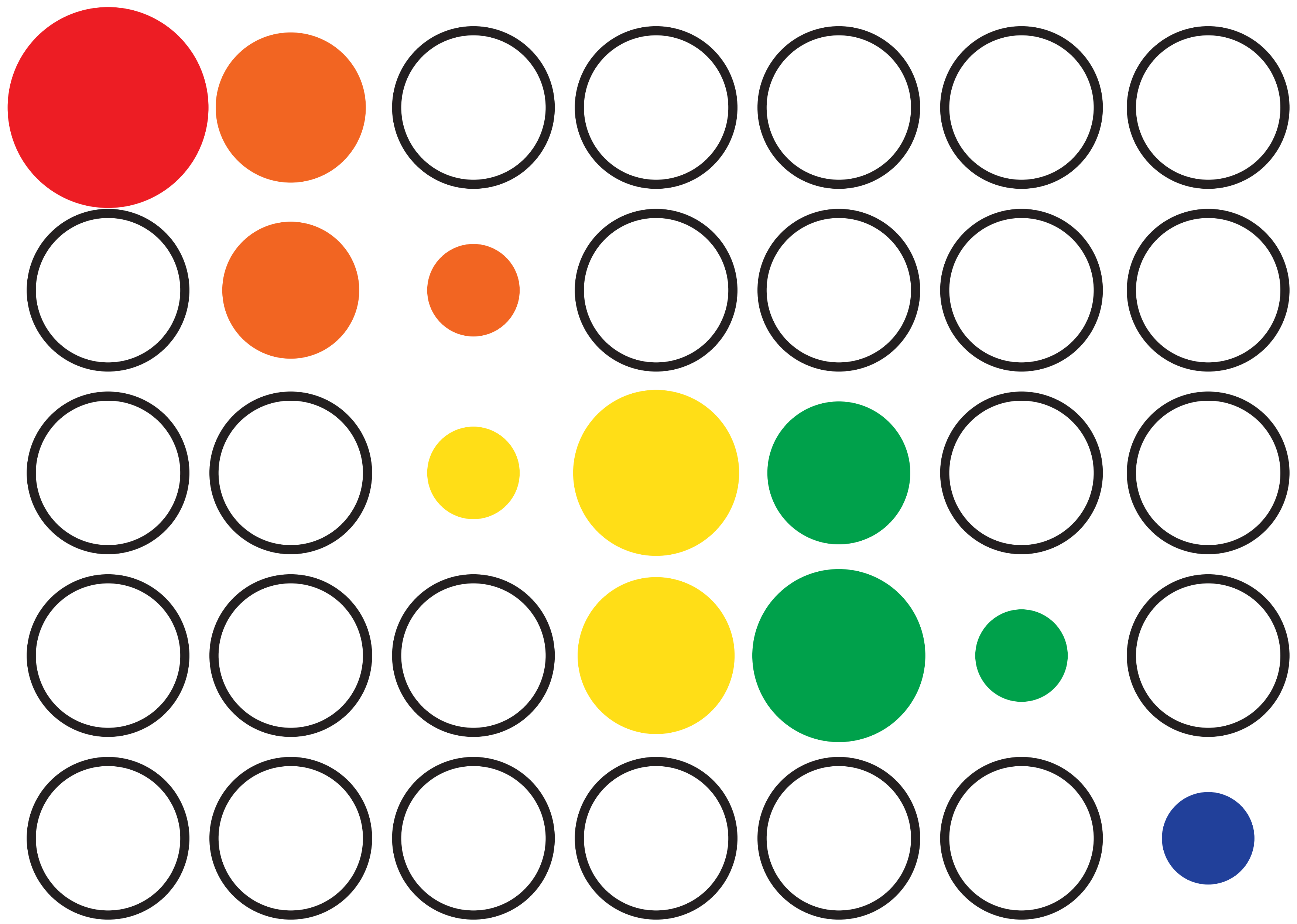
Physics from Light and Time

Cascades ν_e CC | ν_α NC



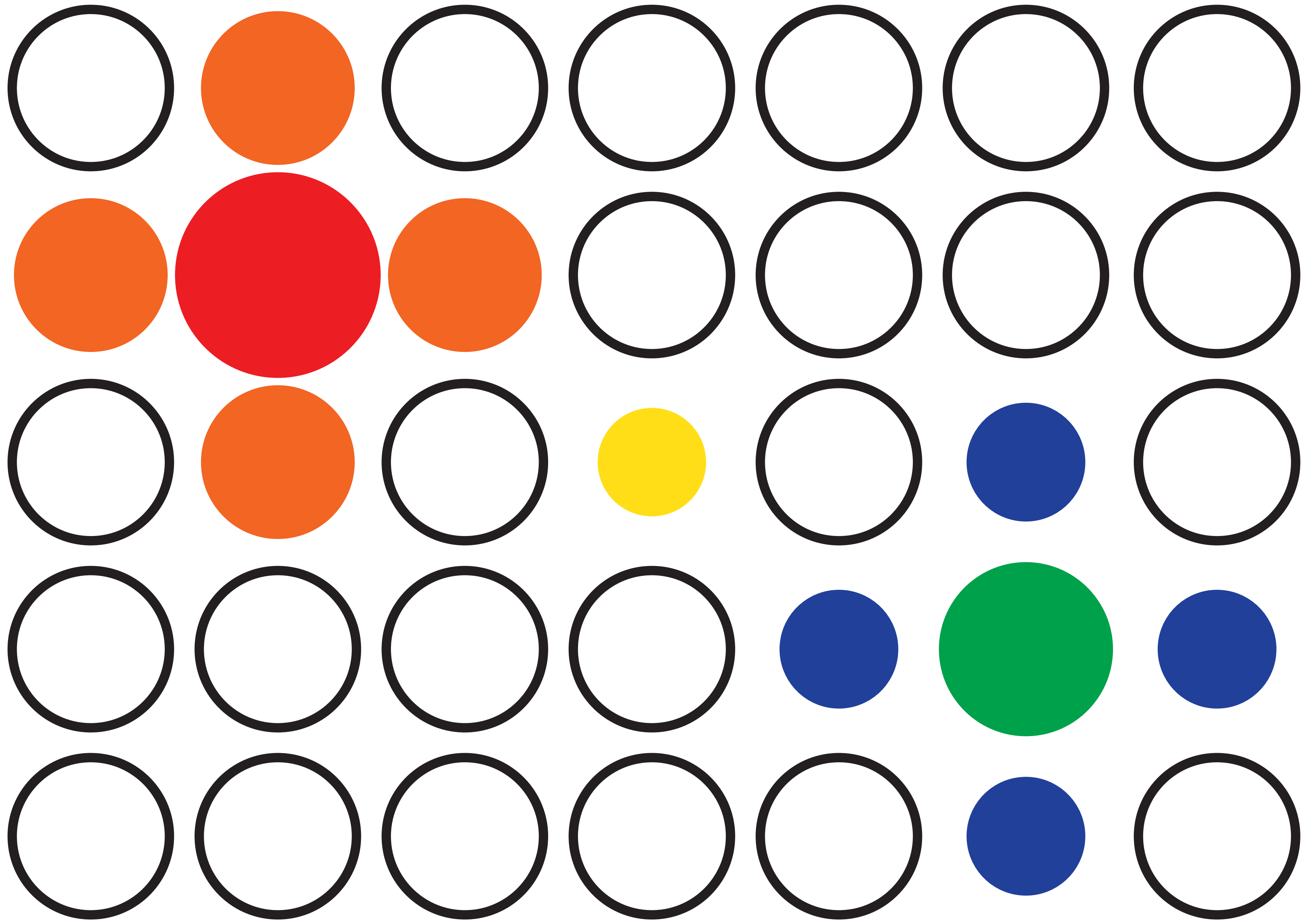
Great energy resolution, but angular reconstruction is challenging

Tracks ν_μ CC



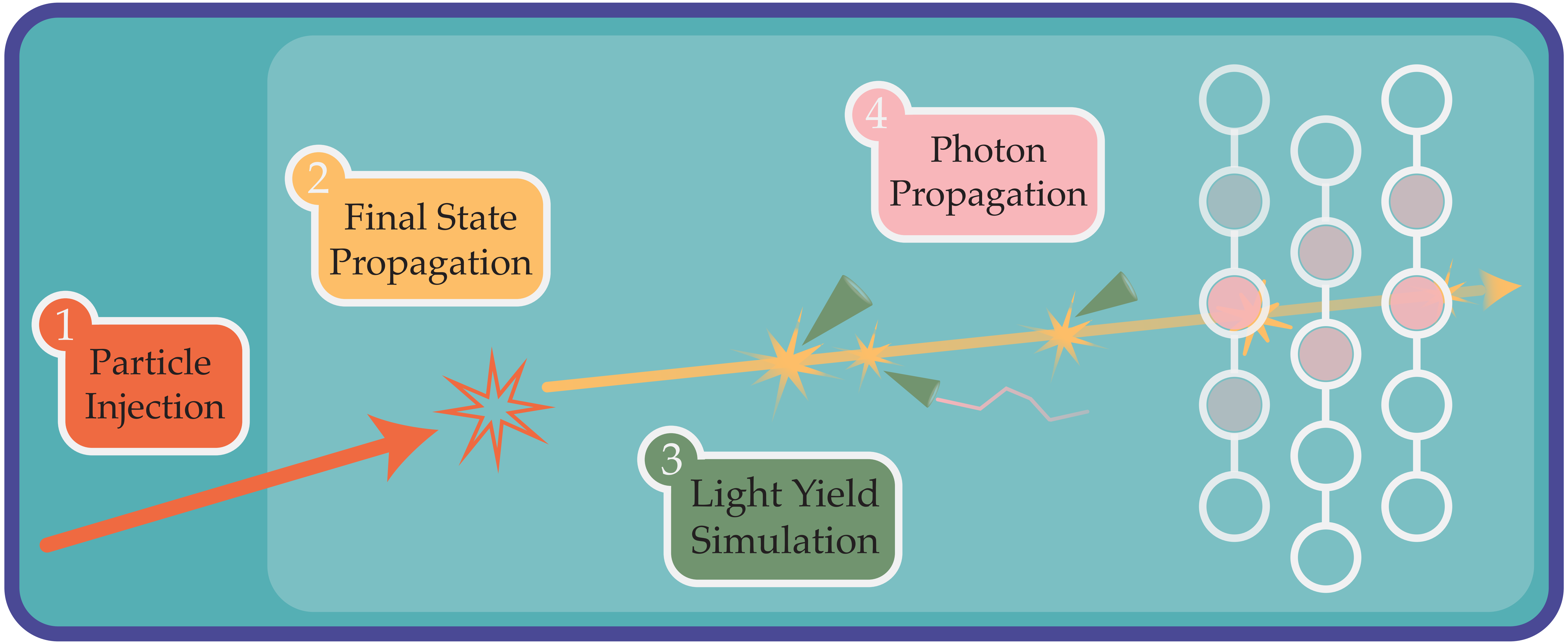
Great directional resolution, but deposited energy not proportional to E_ν

Double bangs ν_τ CC



Signature of ν_τ CC events

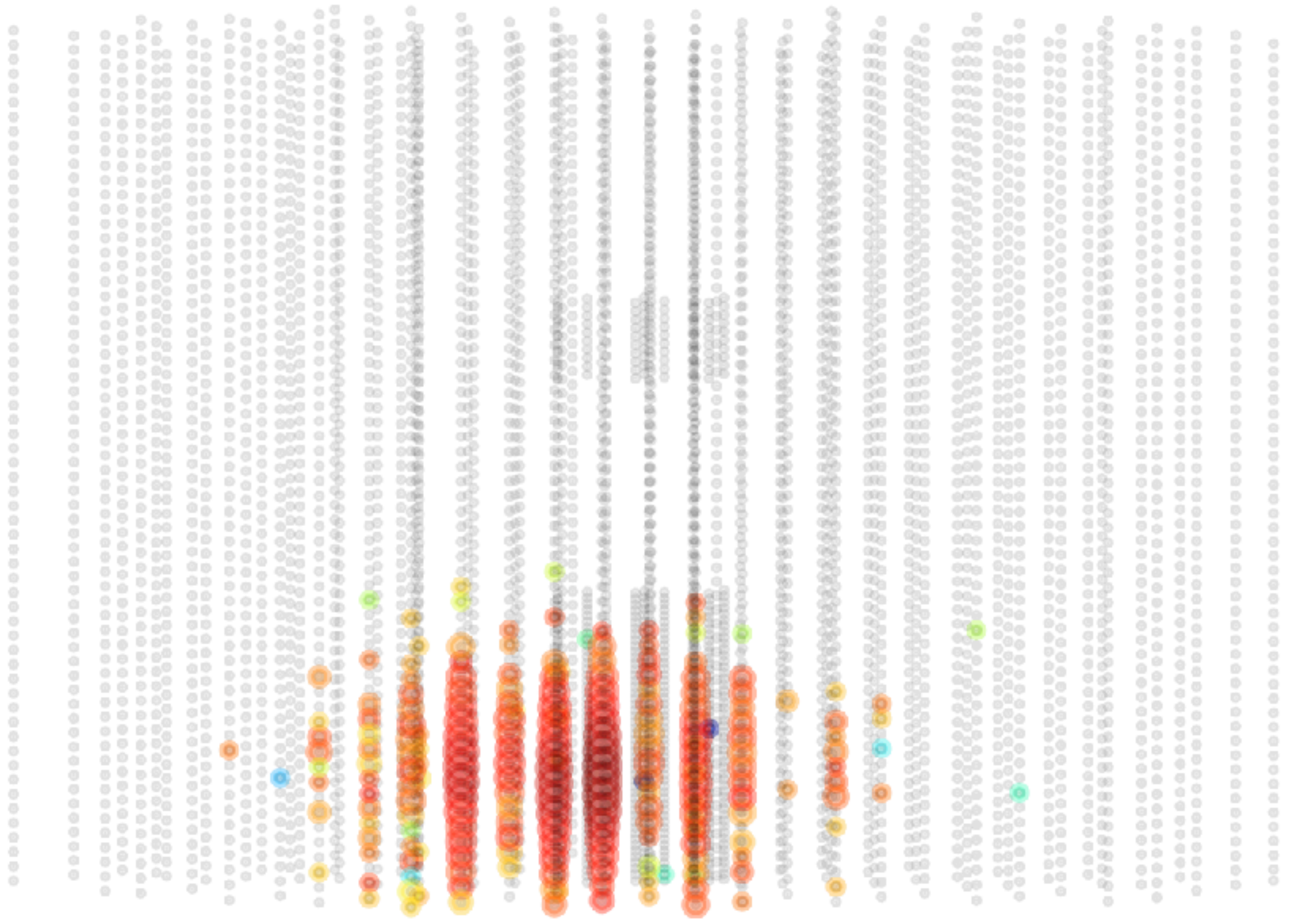
Prometheus Open-Source Simulation Framework



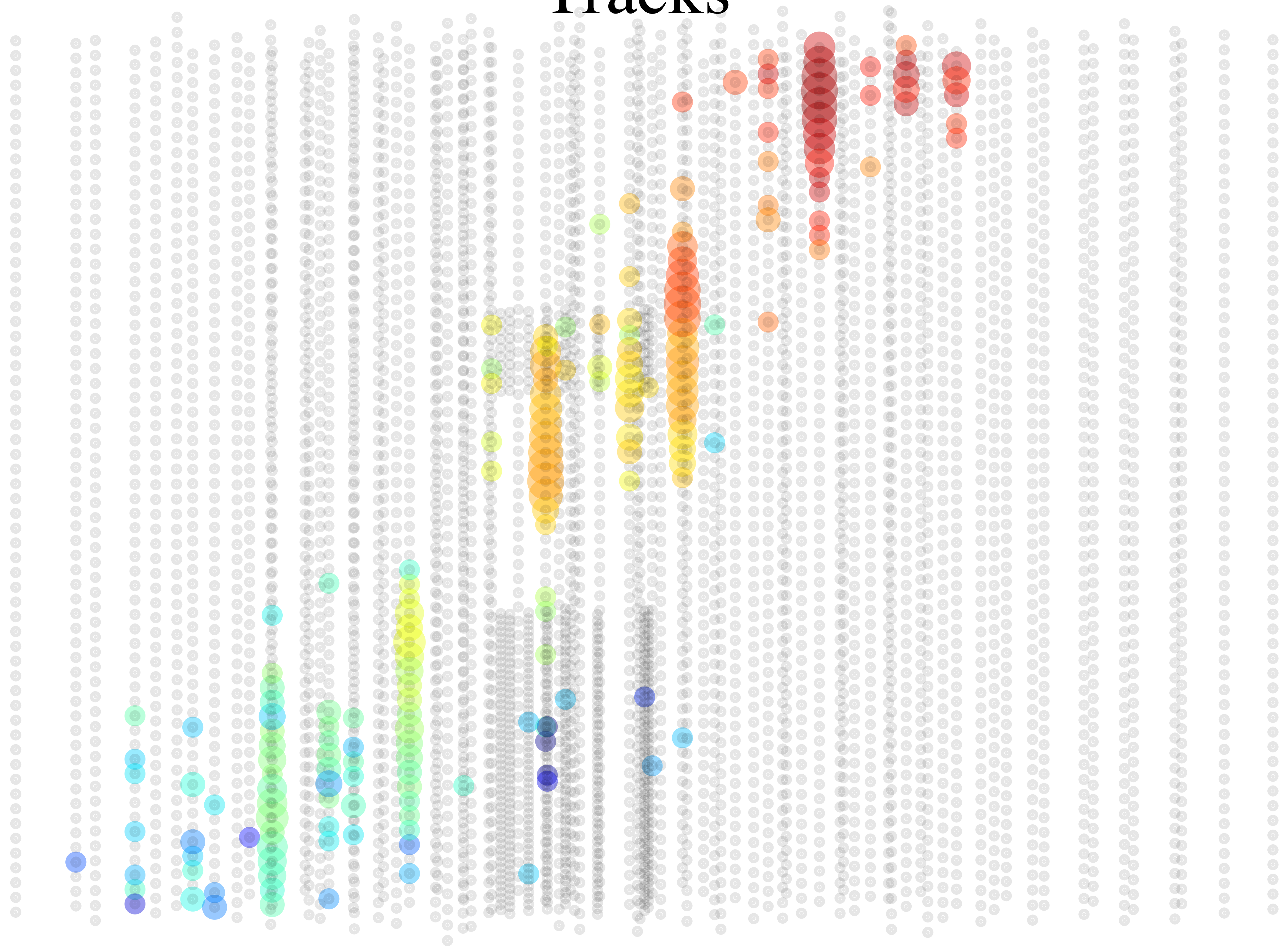
- Prometheus provides support for full simulation chain
- Ice- and water-based detectors
- Photon-level information enabling detailed ML and theoretical studies

In-Ice Event Displays

Cascades

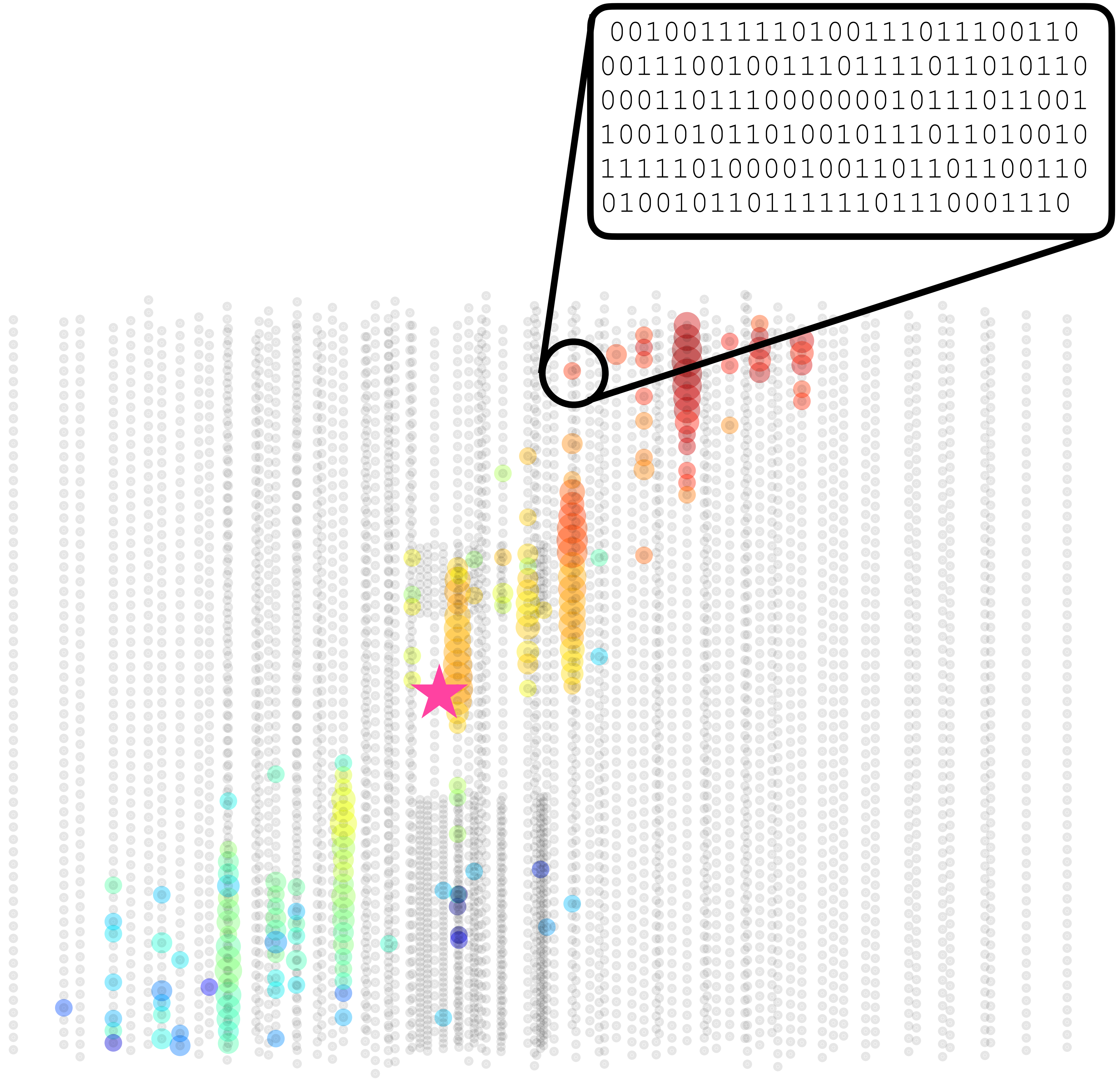


Tracks



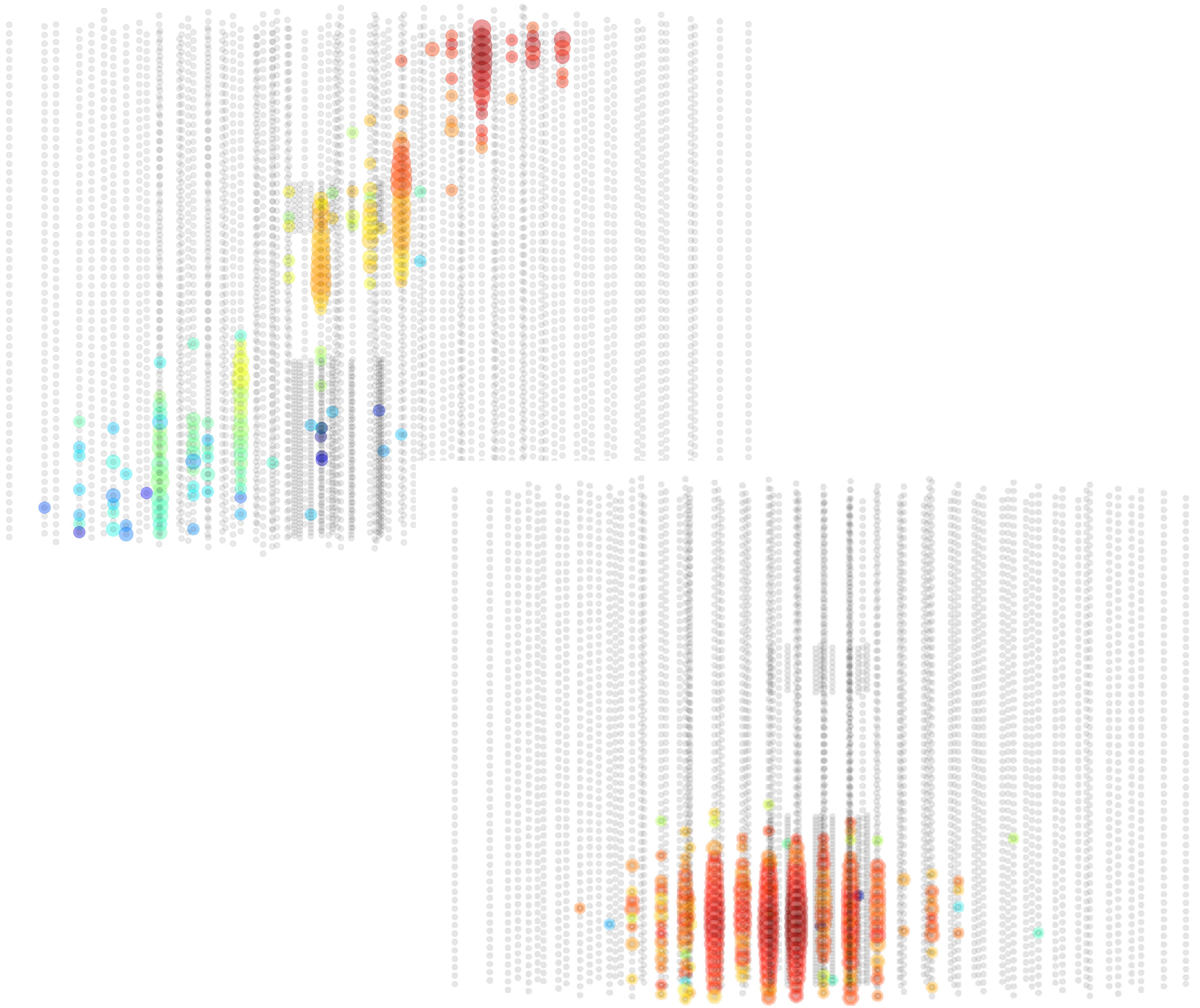
Binaryification

- For each event create a coordinate system centered at the charge-weighted center of gravity
- For each OM compute $\bar{t}$, q_{tot}, and (r, θ, ϕ) and convert to binary via, e.g. Float32
- Concatenate these values !
- With our dataset, we were able to encode each event into 8-qubit systems



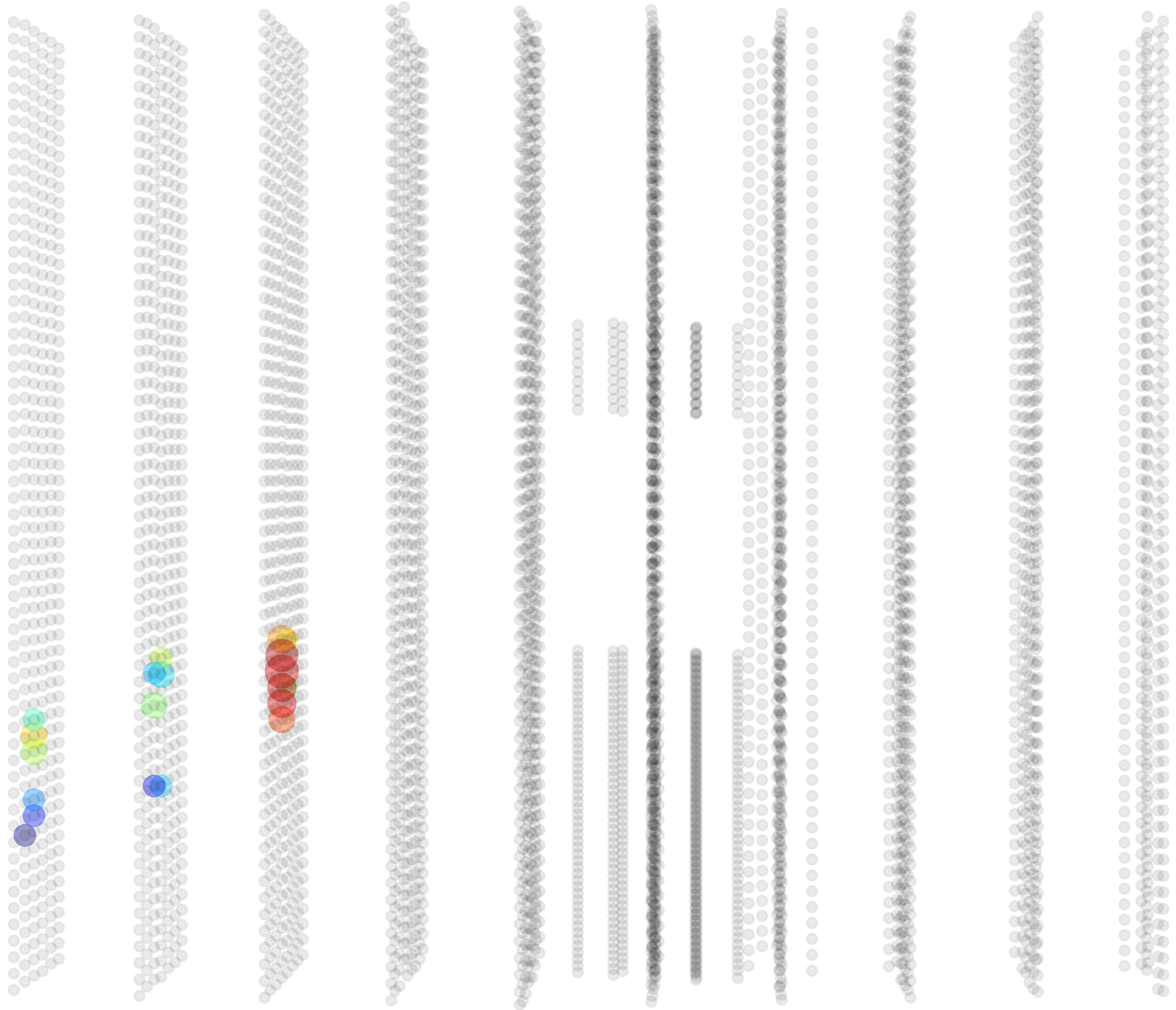
Differentiating Tracks and Cascades

- We wanted to see whether this can be used to analyze physics data
- Compare CDF of polar angle for tracks and cascades
- Expect a more uniform distribution for cascades and peaked for tracks

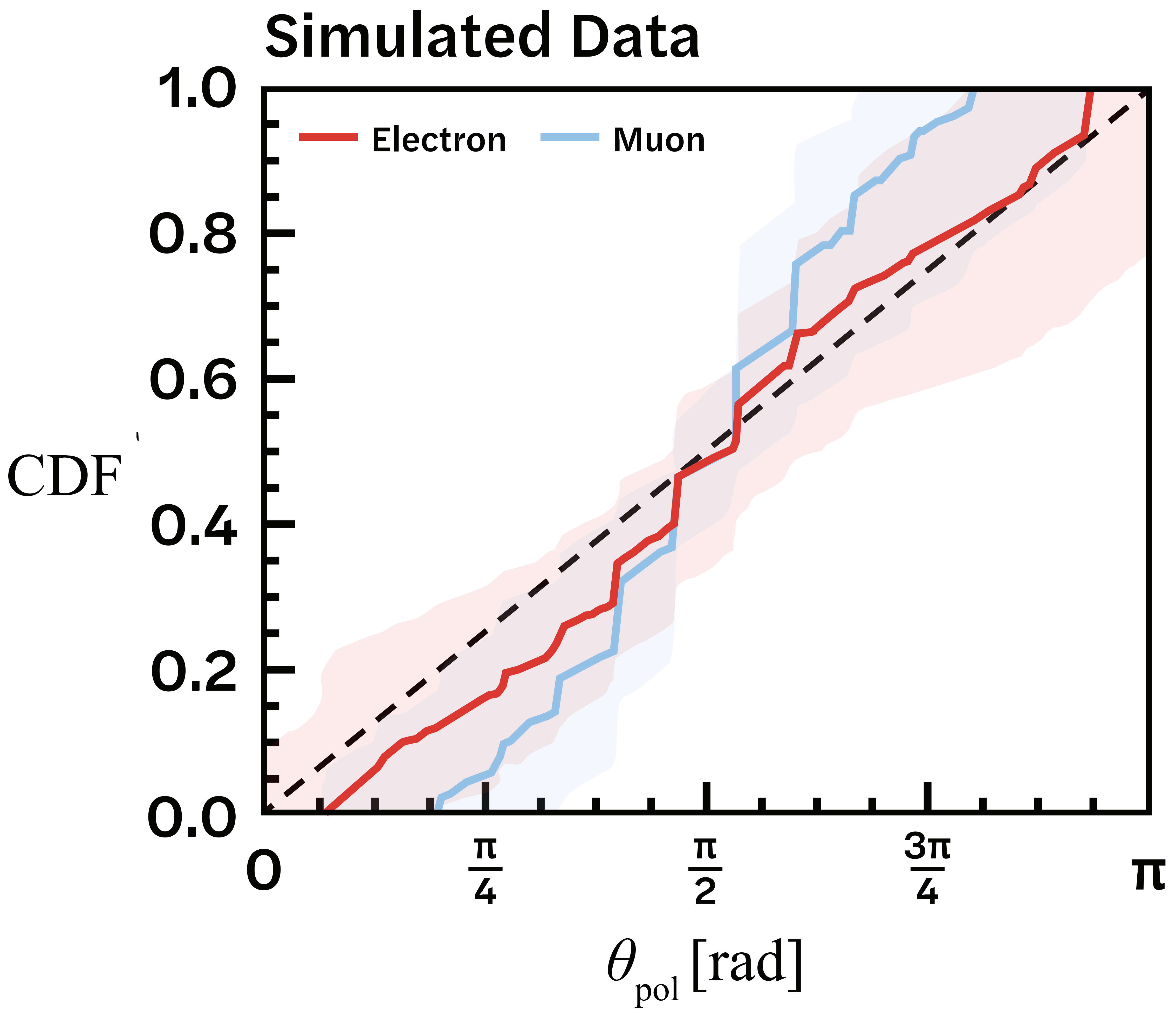


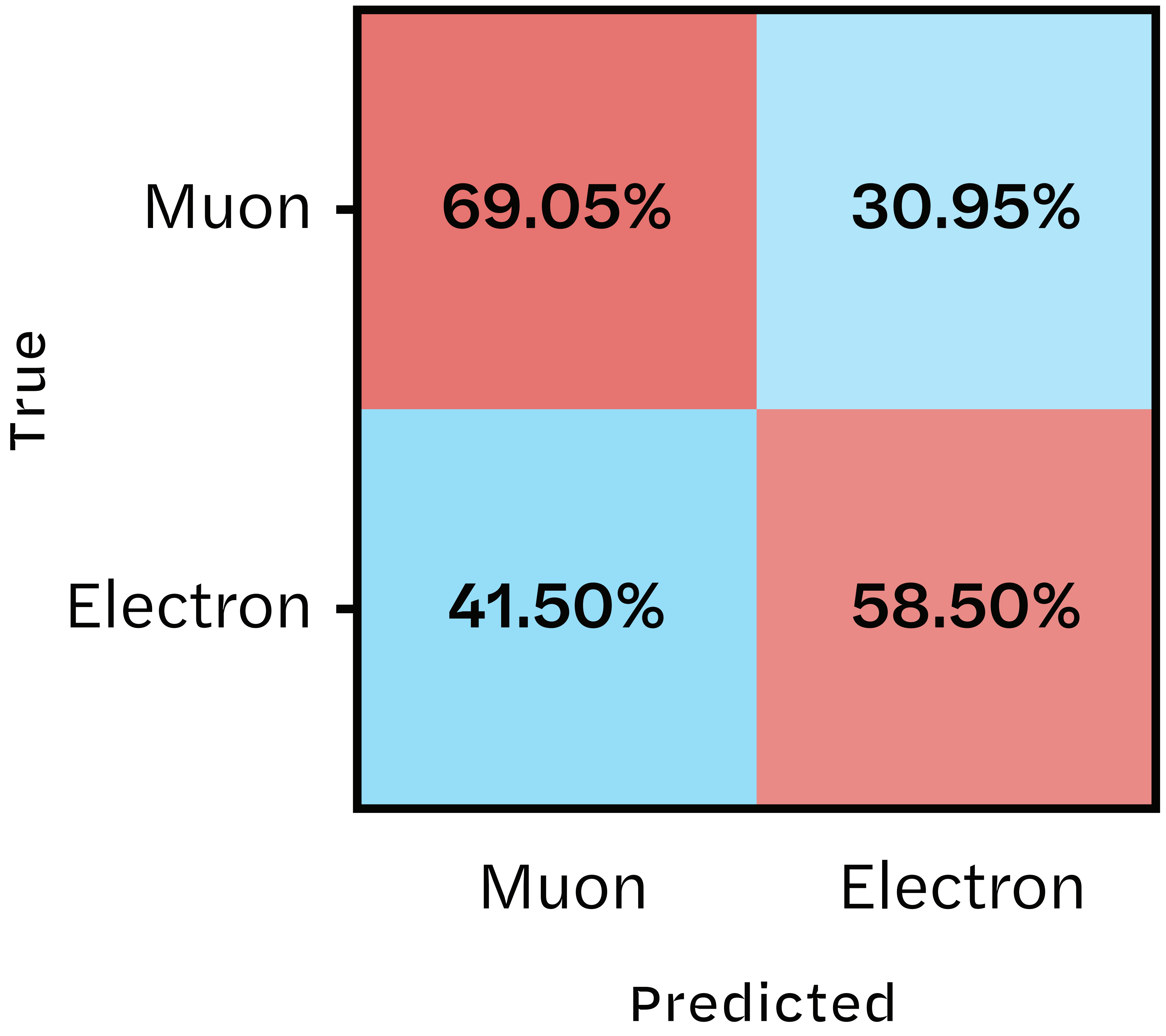
Simulated Dataset

- Restricted ourselves to events that could be encoded in 8 qubits
- Simulated events with energies between 100 GeV and 50,000 GeV
- At least 20 photons recorded and at most 20 OMs triggered



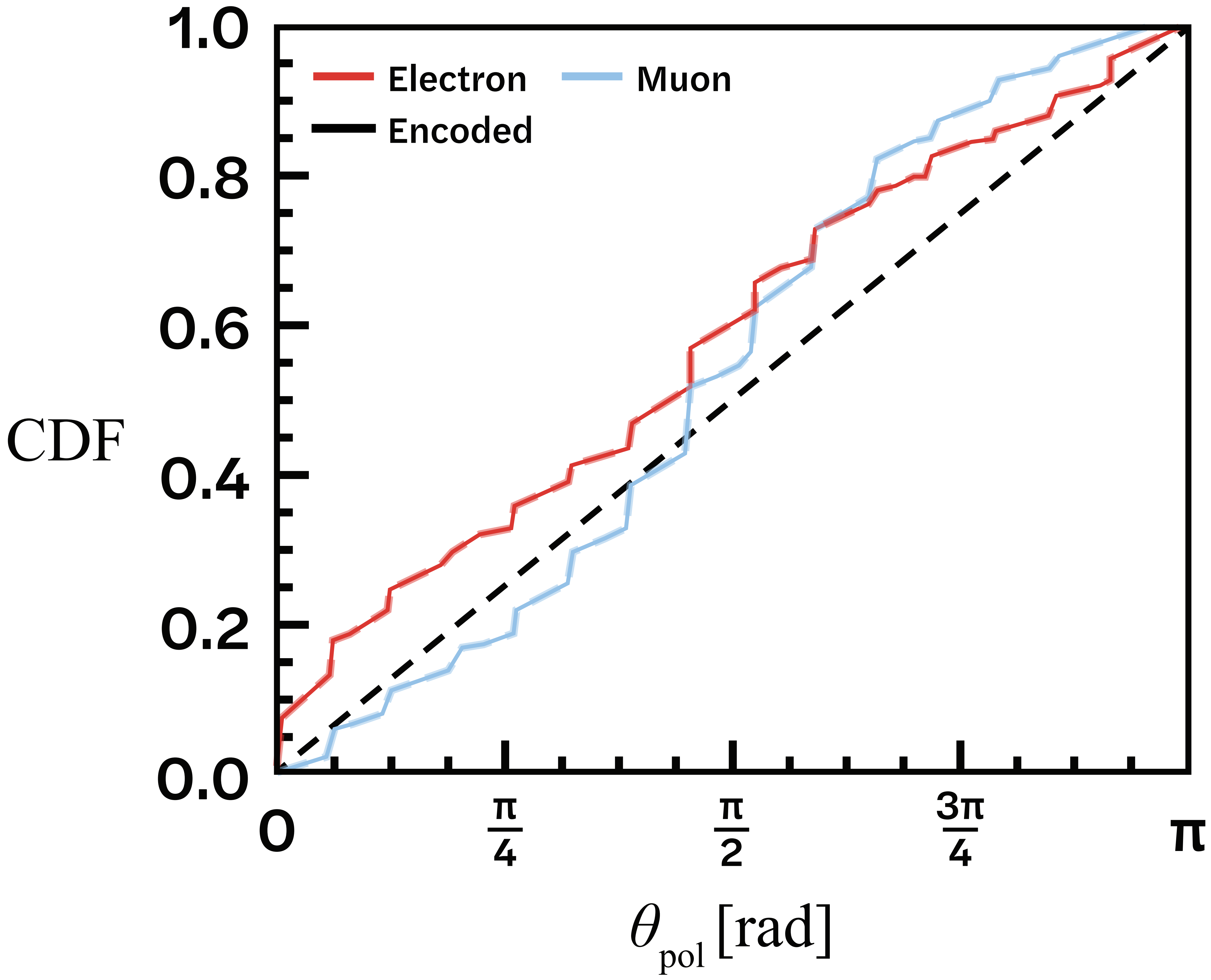
Data Going In





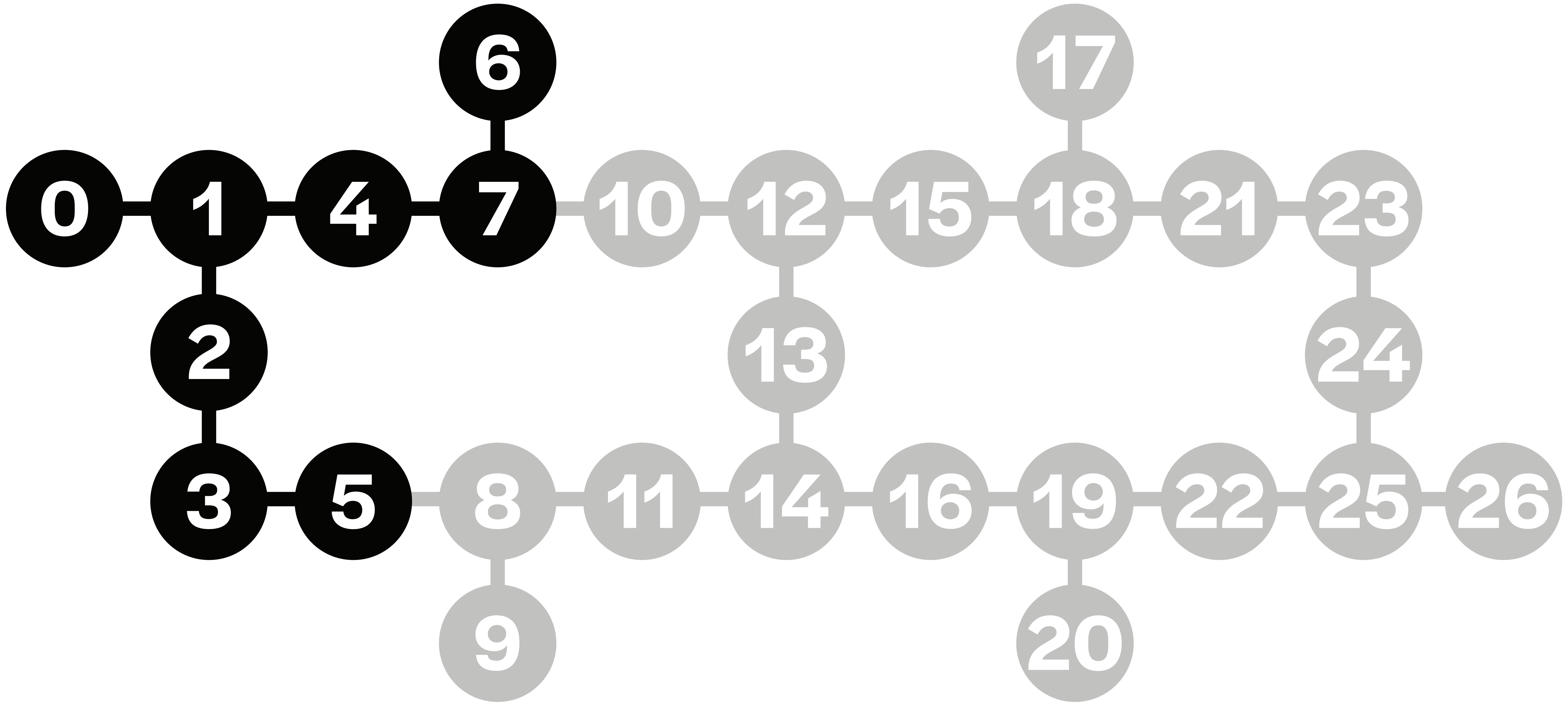
Embedded Data

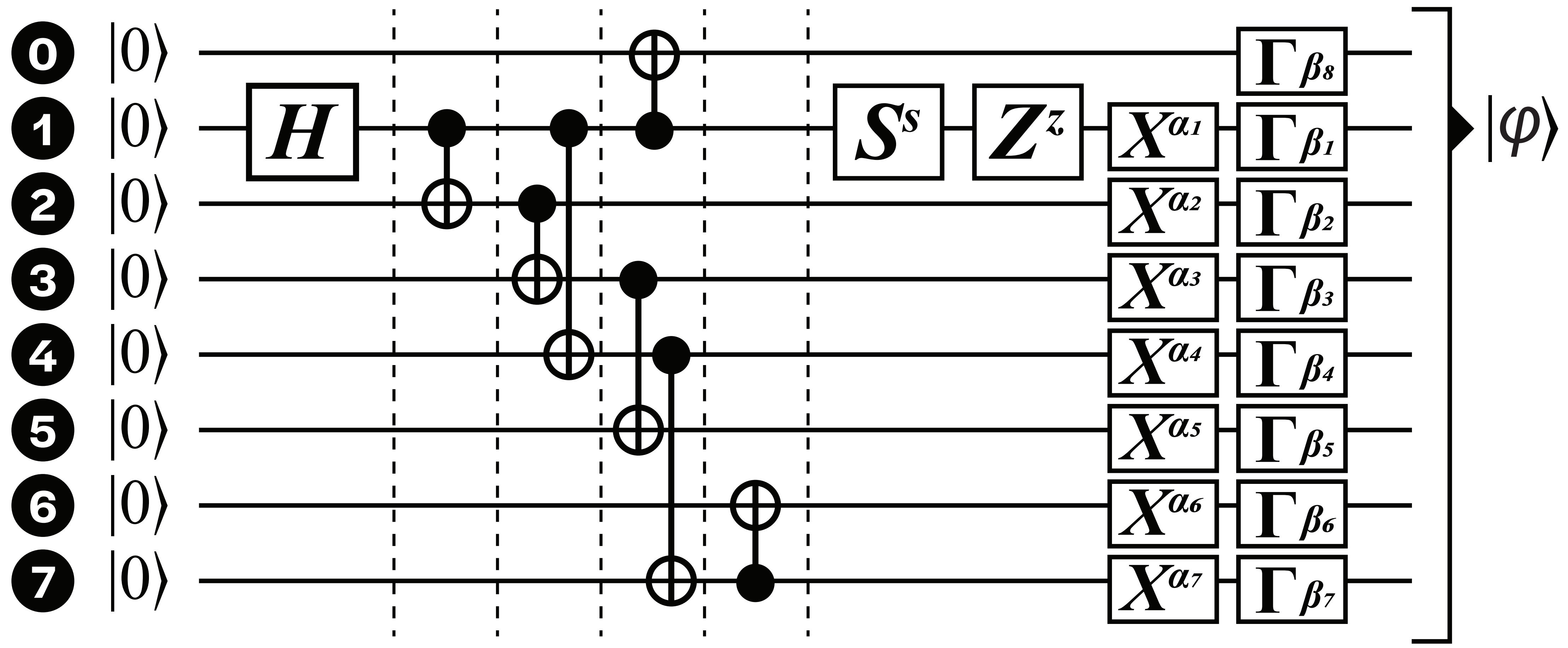
- We encoded our data into 680^{+18}_{-25} 8 qubit states
- The fidelity of the embedding had a fidelity $84.32\,\%^{+0.69\%}_{-1.08\%}$ with respect to the classical data
- Systematic shift upward for both tracks and cascades



IBM Q Cairo Backend

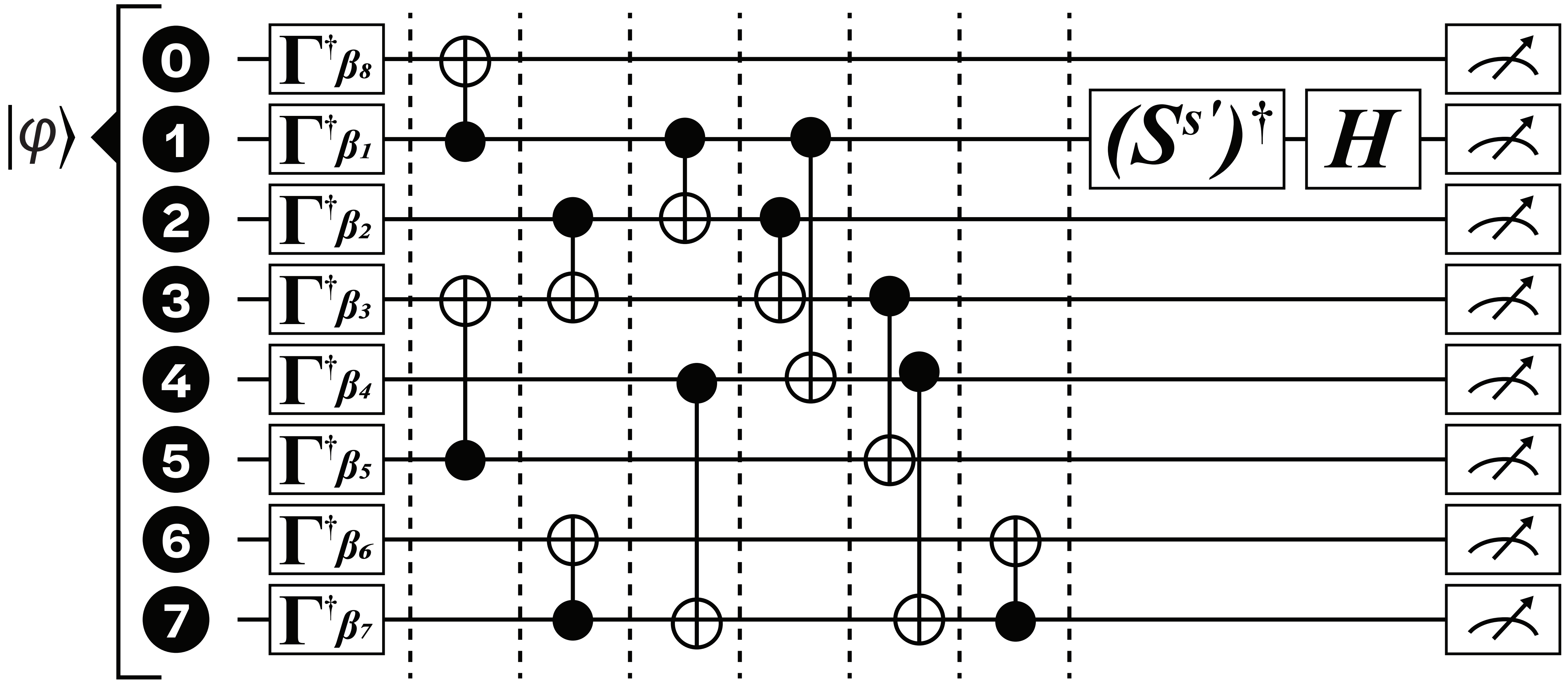
- After running our encoding procedure we embedded the events on the IBMQ Cairo backend
- Modified circuit to maximize parallelization of 2-qubit gates



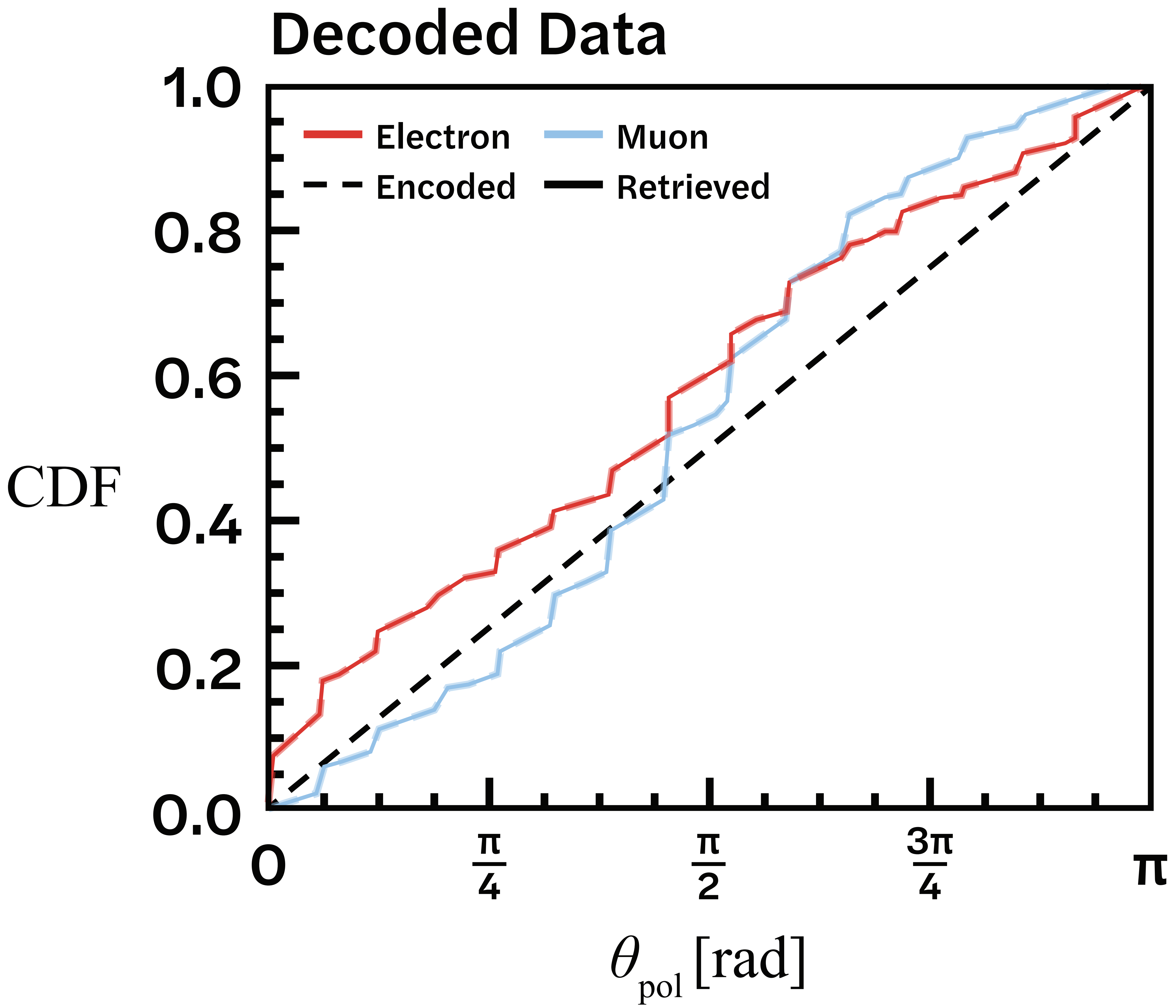


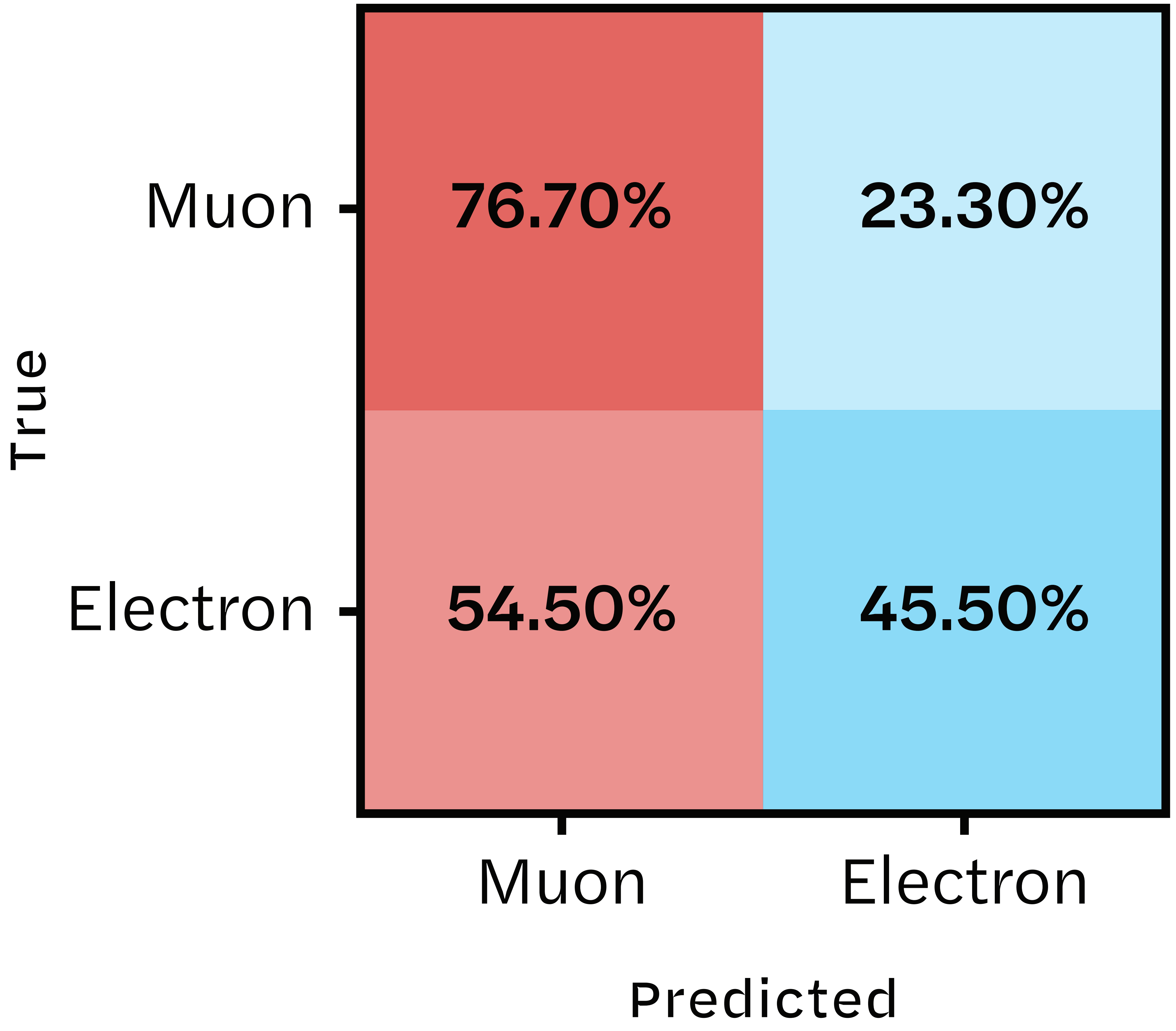
Reading Out the Data

- Finally we read out the data via the decoding circuit
- Again, we optimized the circuit to maximize parallel processing
- We then measure the state of each qubit to reconstruct the initial state



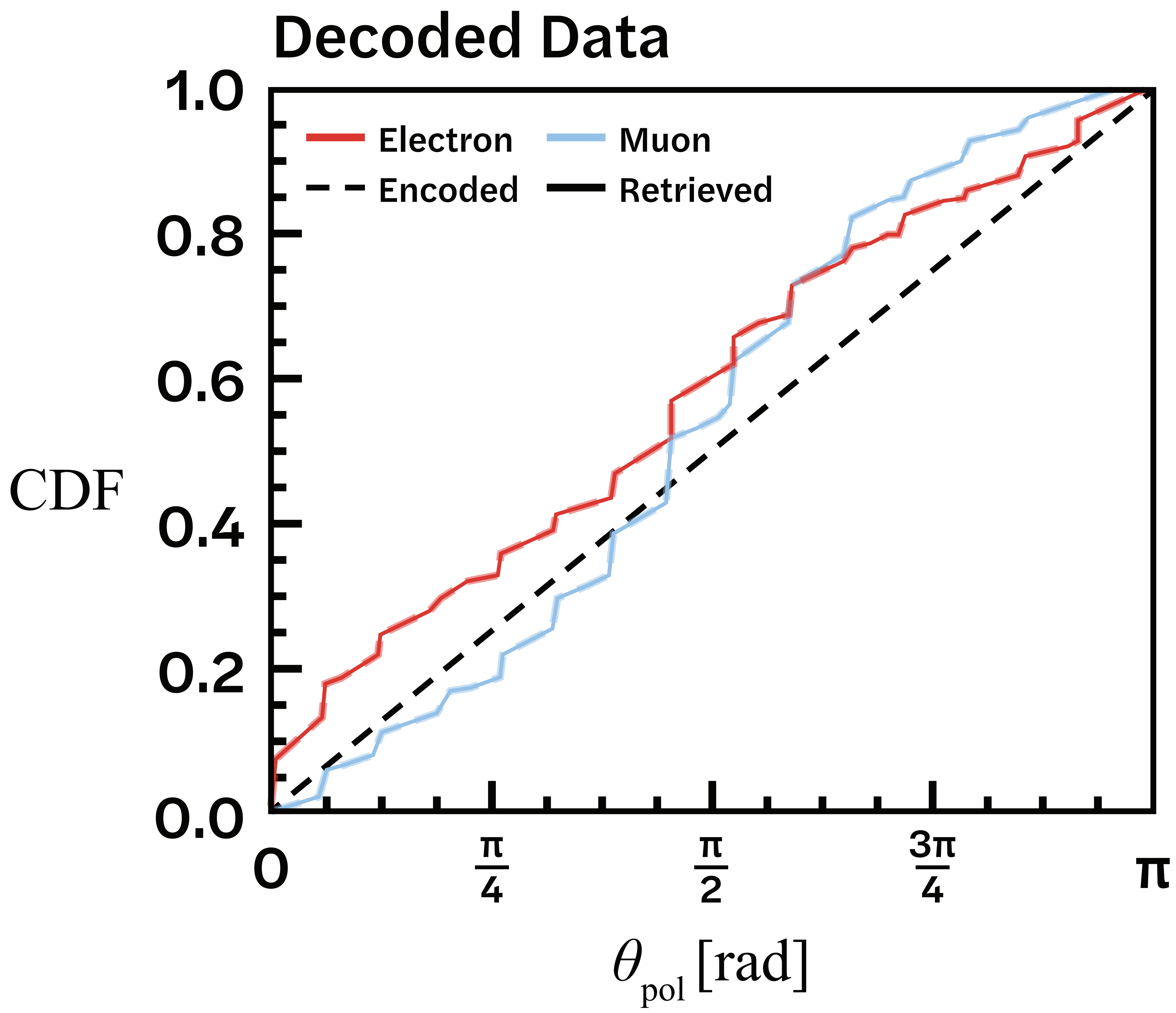
Data Coming Out





Data Coming Out

- Encoded data recovered with $100\,\%^{+0.0\%}_{-1.04\%}$ fidelity
- Discrepancy between true and encoded data made classification fail



Summary Remarks on This Study

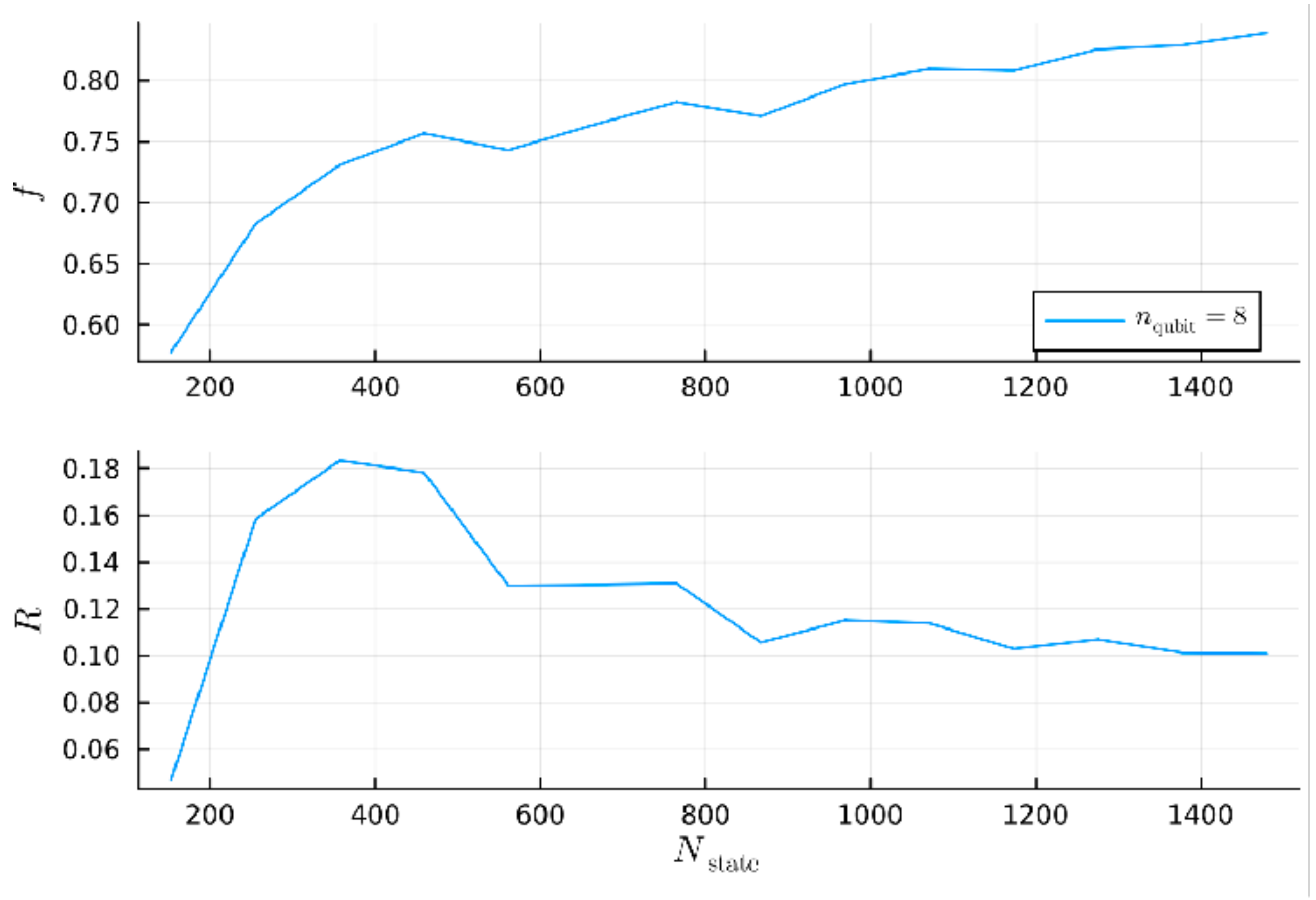
- The high fidelity between encoded and retrieved data shows the embedding protocol is robust to current, noisy quantum computers
- The embedding procedure is not sufficiently faithful to desired data
- No proof whether lack of fidelity is inherent or result of imperfect optimization

Outline

- Encoding Information in Quantum Random Access Codes
- Example application to neutrino telescope data
- **Concluding remarks**

Looking Forward on QRACs in Physics

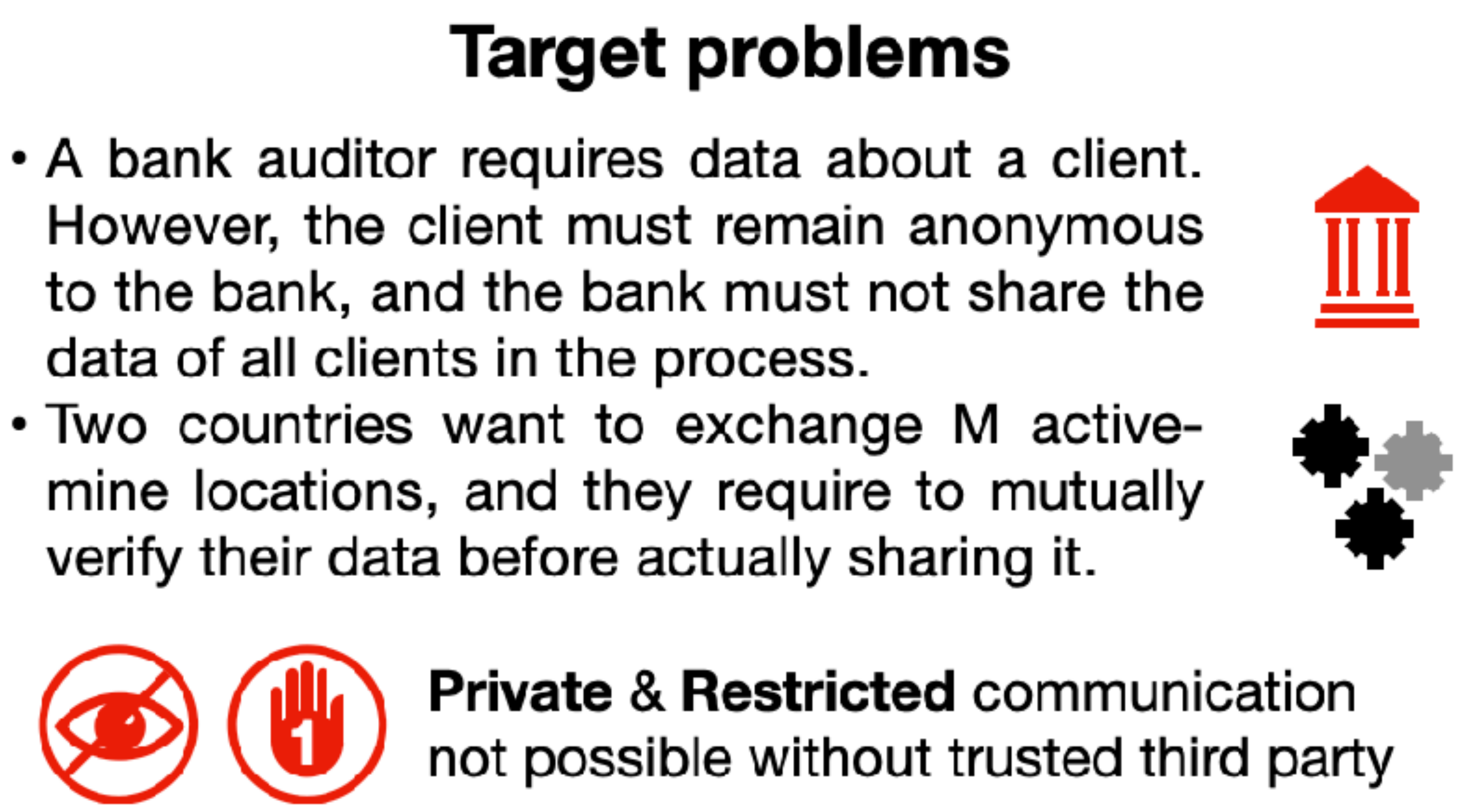
- Understanding whether data can be compressed is imperative for understanding whether QRACs will have physics potential
- Moving data analysis into the quantum circuit, e.g. via quantum VAEs and NNs, should also be explored



$$R = \mathrm{MI}(f) \times \frac{N_{\text{classical}}}{N_{\text{quantum}}}$$

Final Comments about QRACs

- QRACs have interest beyond encoding / compressing data
- We've recently realized potential to use this protocol for private and restricted communication
  - Since information is destroyed as it is read, one can enforce a limit on how much information is known without knowing what information will be read



Conclusions

- Near-term, noisy quantum computers have the potential to aid in high-energy physics
- QRAC protocol can potentially lead to data compression with relatively few qubits, but more studies needed
- Current encoding on 8 qubits does not offer high-enough fidelity to be straightforwardly applied to physics data
- Applications of QRACs exist beyond compression and storage, motivating further study of algorithm's expressive properties

Thank you:-)